

**Tests with Correct Size in the Simultaneous  
Equations Model**

Copyright 2002

by

Marcelo Jovita Moreira

## Abstract

Tests with Correct Size in the Simultaneous Equations Model

by

Marcelo Jovita Moreira

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Thomas Rothenberg, Chair

Classical statistical theory is employed to find tests for structural parameters with correct size even when identification is weak. The family of exactly similar tests is characterized in the limited-information model where the reduced-form errors are normal with known covariance matrix. A version of the score test and a particular two-step procedure based on a preliminary test for identification are shown to be members of this family of similar tests. In addition, a power bound is derived for the family. The Anderson-Rubin test is shown to be optimal when the model is just identified; no test is uniformly most powerful when the model is overidentified.

Again assuming normal reduced form errors with known covariance matrix, a general method is proposed for finding a similar test based on the conditional distribution

of an arbitrary test statistic. The method is applied to the two-stage least-squares Wald statistic and to the likelihood ratio test statistic. Dropping the assumption of known error distribution, it is found that slightly modified versions of these conditional tests are similar under weak-instrument asymptotics. Monte Carlo simulations suggest that the conditional likelihood ratio test is essentially optimal when identification is strong and also dominates other similar tests when identification is weak.

---

Professor Thomas Rothenberg  
Dissertation Committee Chair

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1	An Example . . . . .	3
2	The Model . . . . .	6
3	Testing with Weak Instruments . . . . .	8
4	Classical Testing Theory . . . . .	12
<b>2</b>	<b>Tests with Correct Size when Instruments May Be Arbitrarily Weak</b>	<b>15</b>
1	The Family of Similar Tests . . . . .	16
1.1	Similar Tests Based on Pivotal Statistics . . . . .	18
1.2	Pre-testing Procedures . . . . .	21
2	Power Functions . . . . .	22
3	Score Test . . . . .	24
4	Conclusions . . . . .	26
<b>3</b>	<b>A Conditional Likelihood Ratio Test for Structural Models</b>	<b>27</b>
1	Similar Tests Based on Nonpivotal Statistics . . . . .	28
1.1	Examples of Conditional Tests . . . . .	29
2	The Conditional Likelihood Ratio Test . . . . .	32
3	Monte Carlo Simulations . . . . .	36
4	Conclusions . . . . .	37
<b>4</b>	<b>Approximations when the Error Distribution is Unknown</b>	<b>39</b>
1	Conditional Approach Revisited . . . . .	40
1.1	The Anderson-Rubin Test . . . . .	40
1.2	The Score Test . . . . .	41
1.3	The Wald Test . . . . .	43
1.4	The Likelihood-Ratio Test . . . . .	43
2	Weak Instrument Asymptotics . . . . .	44
3	Monte Carlo Simulations . . . . .	47
4	Confidence Regions . . . . .	49

5	Extensions . . . . .	50
6	Conclusions . . . . .	52

# Chapter 1

## Introduction

Applied researchers are often interested in making inferences about the coefficients of endogenous variables in a linear structural equation. Identification is achieved by assuming the existence of instrumental variables uncorrelated with the structural error but correlated with the endogenous regressors. If the instruments are strongly correlated with the regressors, standard asymptotic theory can be employed to develop reliable inference methods. However, as emphasized in recent work by Bound, Jaeger and Baker (1995), Dufour (1997), and Staiger and Stock (1997), these methods are not satisfactory when instruments are only weakly correlated with the regressors. In particular, the usual tests and confidence regions do not have correct size in the weak instrument case.

In this thesis, classical statistical theory is employed to find tests with correct size even when identification is weak. When the reduced-form errors are normal

with known covariance matrix, we describe a large family of tests that are exactly similar—that is, they have same null rejection probability for all values of the nuisance parameters. Dropping the assumption of known error distribution, we find that slightly modified versions of these tests are approximately similar in moderate sized samples. Monte Carlo simulations suggest that one member of the family based on the likelihood ratio statistic is essentially optimal when identification is strong and also dominates other similar tests when identification is weak.

In the remaining sections of this chapter, the linear limited-information simultaneous equations model is described and the econometrics literature on weak instruments surveyed. In addition, a few key concepts from classical testing theory are summarized. In Chapter 2, exponential family theory is employed to characterize the class of similar tests in a two equation model with reduced-form disturbances that are normal with known covariance matrix. When the model is overidentified, there exists no uniformly most powerful test; the (upper bound) power envelope for the family of similar tests is computed. In Chapter 3, a simple general method is proposed for finding similar tests starting with an arbitrary test statistic. The method employs a critical value function derived from the conditional distribution of the test statistic given a subset of the sufficient statistics. The conditional test based on the likelihood ratio statistic is shown to have good power properties. Finally, in Chapter 4 the results are generalized to the case where there are more than two endogenous variables and where the errors distribution is unknown.

## 1 An Example

An interesting example of structural inference with weak instruments is a recent paper by Angrist and Kreuger (1991) on the returns to education. It would be expected that two people, similar in native ability but differing in their educational attainment, will be treated differently in the job market. Education provides some skills valued by firms and hence the better educated person is likely to earn a higher wage. To estimate the returns to education from census data, economists sometimes postulate a linear stochastic equation relating log weekly earnings to years of education, with additional covariates controlling for other features that make one person different from another. The error term represents the effects of person-to-person differences that have not been controlled for.

Of course, these data do not result from a controlled experiment where education is randomly assigned to people. Years of education is, to a large extent, chosen by the individual. Hence both earnings and education must be treated as endogenous variables. If some of the factors that influence educational choice are also factors that constitute the error term, education will be correlated with the error term and a least squares regression of log earnings on years of education will not yield an estimate of the true returns to education. In practice, the explanatory variables included in the model do not capture much of the variation of earnings. Thus there is the potential for considerable “simultaneous equations” bias. However, if we have data on variables that explain the variation in years of schooling but these variables do not directly

effect earnings potential, then these variables can be used as instruments to estimate the return to education.

Angrist and Krueger (1991) propose to estimate the effect of education on earnings using as instruments dummy variables that indicate in which quarter the individual was born. They argue that quarters-of-birth are an exogenous source of variation in educational attainment, due to existing compulsory schooling laws. In many states, the law requires children to enter school in the year of their sixth birthday, and to study until the age of sixteen. Consequently, children born in the first months of the year start school at a later age and may drop out before finishing the tenth grade. On the other hand, children born later in the year start school at about six years old and finish the tenth grade before reaching the legal drop-out age. Thus, people born in the last quarter are more likely to have a higher level of education than people born in the first quarter of the year.

The association between quarter-of-birth and both age at school entry and educational attainment is supported by two findings. First, the effect of birth on school attainment varies across states, depending on the legal drop-out age. The drop-out rate is considerably higher in states with a sixteen year age requirement compared to states with a longer requirement. Second, the relationship is weaker for more recent cohorts. Since the average education level has increased over time, the laws are less likely to influence more recent cohorts.

To estimate returns to schooling using quarters-of-birth as instruments, Angrist

and Krueger (1991) mainly use the U.S. Census data from 1980. Their sample consists of men born in the United States between 1930 and 1959 who reported age, sex, race, quarter of birth, weeks worked, years of schooling and salary. They then estimate wage equations considering different specifications that combine different sets of instruments (quarters-of-birth possibly interacting with years-of-birth and regions-of-birth) and sets of covariates (race, metropolitan area, marital status, age, and age squared). In a typical case, they have about 350,000 observations and up to 178 instrumental variables.

Bound, Jaeger and Baker (1995) replicate Angrist and Krueger's results for the 1930-39 cohort and question the reliability of quarters-of-birth as instruments for educational attainment. Their criticism is twofold. First, quarters-of-birth can be correlated with unobserved characteristics of the individual. Therefore, these instruments are not truly exogenous, being correlated with earnings after controlling for education. This question about the exogeneity of the instruments is addressed by Angrist and Krueger (1991) and will not be considered here.

The second (and for our purposes more relevant) criticism is that quarter-of-birth is only weakly correlated to educational attainment in some specifications (mostly those that include age and age squared as covariates). This argument is supported by the low values of the F-statistic that tests the statistical significance of the instruments' coefficients. Although identification seems weak, the reported standard errors on the estimates are surprisingly small. To illustrate that these standard errors are

unreliable indicators of the accuracy of the estimators, Bound, Jaeger and Baker (1995) conduct a simulation experiment where they estimate returns to schooling using randomly generated instruments that have no correlation at all with education. The results are striking since the standard errors are similar to those reported by Angrist and Krueger (1991). Bound, Jaeger and Baker conclude that the Angrist-Krueger confidence intervals are unreliable since results similar to theirs would have been obtained in situations where the true confidence interval would have to be the entire real line. Standard asymptotic approximations, they argue, can give very misleading information when the correlation between the instrument and endogenous variable is weak even for the large sample sizes available in the U.S. Census. The need for more reliable inference methods in the weak instrument case motivates the research that follows.

## 2 The Model

To emphasize the main ideas and avoid tedious notation, we will consider in Chapters 1-3 a simple structural equation containing only one explanatory endogenous variable without additional covariates. Extensions to the general case of many explanatory variables (both endogenous and exogenous) will be developed in Chapter 4. Let  $y_1$  and  $y_2$  be  $n \times 1$  vectors of observations on two endogenous variables related

by the structural equation

$$(1) \quad y_1 = \beta y_2 + u$$

where  $u$  is an  $n \times 1$  unobserved error vector and  $\beta$  is an unknown scalar parameter. We assume that the elements of  $u$  are independent and identically distributed (i.i.d.) with mean zero and variance  $\sigma^2$ . This equation is assumed to be part of a larger linear simultaneous equations system which implies that  $y_2$  is correlated with  $u$ . However, the complete system contains exogenous variables which can be used as instruments for conducting inference on  $\beta$ . In the spirit of limited information analysis, we shall not specify the remaining structural equations but simply assume that they can be solved so that  $Y = [y_1, y_2]$  can be expressed as a set of reduced-form regression equations

$$(2) \quad \begin{aligned} y_1 &= Z\pi\beta + v_1 \\ y_2 &= Z\pi + v_2 \end{aligned}$$

where  $Z$  is an  $n \times k$  matrix of nonrandom exogenous variables having full column rank  $k$ ;  $\pi$  is a  $k \times 1$  vector and the  $n$  rows of the  $n \times 2$  matrix of reduced form errors  $V = (v_1, v_2)$  are i.i.d. with mean zero and  $2 \times 2$  covariance matrix

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}.$$

The restriction that the regression coefficient in the first reduced-form equation is a scalar multiple of the coefficient in the second equation is implied by the identifying assumption that the exogenous variables do not appear in (1).

The goal is to test the hypothesis  $H_0 : \beta = \beta_0$ , treating  $\pi$  and  $\Omega$  as nuisance parameters. A test is said to be of size  $\alpha$  if the probability of rejecting the null hypothesis when it is true does not exceed  $\alpha$ . That is:

$$\text{prob}(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}) \leq \alpha.$$

for all admissible values of  $\pi$  and  $\Omega$ . Since these parameters are unknown, finding a test with correct size is nontrivial. The task is simplified if one can find tests whose null rejection probability does not depend on the nuisance parameters at all. These tests are called *similar tests*. If, for example, one rejects if some test statistic  $\mathcal{T}$  is greater than a given constant, the test will be similar if the distribution of  $\mathcal{T}$  under the null hypothesis does not depend on  $\pi$ . Such test statistics are said to be *pivotal*. If  $\mathcal{T}$  has null distribution depending on  $\pi$  but it can be bounded by a pivotal statistic, then  $\mathcal{T}$  is said to be *boundedly pivotal*.

### 3 Testing with Weak Instruments

For any matrix  $Q$  with full column rank, let  $N_Q = Q(Q'Q)^{-1}Q'$  and  $M_Q = I - N_Q$ . A commonly used estimator for  $\beta$  is the two-stage least squares estimator  $b_{2SLS} =$

$(y_2'N_Zy_2)^{-1}y_2'N_Zy_1$ . The standardized estimator  $\sqrt{\pi'Z'Z\pi}(b_{2SLS} - \beta)/\sigma$  converges to a standard normal random variable if the elements of  $Z$  are uniformly bounded as  $n$  tends to infinity and  $\pi'Z'Z\pi/n$  tends to a positive number. Under these assumptions,  $\text{plim}(y_2'N_Zy_2 - \pi'Z'Z\pi)/n = 0$ . For some consistent estimator  $\tilde{\sigma}^2$ , a natural two-tailed test of  $H_0$  is to reject if the so-called Wald statistic

$$W = \frac{(b_{2SLS} - \beta_0)^2 y_2'N_Zy_2}{\tilde{\sigma}^2}$$

is larger than  $c_\alpha$ , the  $1-\alpha$  quantile of a chi-square distribution with one degree of freedom. This test is asymptotically similar since

$$\lim_{n \rightarrow \infty} \text{prob}(W > c_\alpha) = \alpha.$$

under the null hypothesis. Thus, for any given  $\pi \neq 0$ , the null rejection probability will be close to  $\alpha$  for a large enough sample size. Unfortunately, since the convergence is not uniform in  $\pi$ , the needed sample size may be extremely large in the weak instrument case where  $\pi'Z'Z\pi/\omega_{22}$  is small. In practice the size of the Wald test may differ substantially from the size based on the asymptotic distribution when identification is weak. In fact, based on earlier work by Gleser and Hwang (1987), Dufour (1997) shows that the true levels of the usual *Wald*-type tests can deviate arbitrarily from their nominal levels if  $\pi$  cannot be bounded away from the origin. This conclusion is reinforced by the analysis of Staiger and Stock (1997) who consider

an alternative asymptotic framework where  $\pi$  tends to the origin as  $n$  tends to infinity. Under this weak-instrument asymptotics, the null rejection probability of the Wald test depends on unknown parameters and is no longer asymptotically similar.

Since weak instruments are often encountered in applied research, it would be desirable to find tests with approximately correct size  $\alpha$  even when  $\pi$  cannot be bounded away from the origin. Some recent research in this direction is surveyed below.

- One possibility is to replace the two stage least squares estimator by the maximum likelihood estimator in forming the Wald statistic. Alternatively, one could use another test statistic such as the likelihood ratio. However, as shown by Wang and Zivot (1998), the size problem holds for this test also.
- For some testing problems, improved critical values can be obtained by using a higher-order Edgeworth expansion of the distribution function of the test statistic. This approach is described in Rothenberg (1984). Unfortunately, the lack of uniform convergence in the first-order asymptotics carries over to higher order asymptotics and the *Dufour critique* remains valid for the “improved” critical values.
- Instead of using the limiting distribution of  $W$  to find a critical value one could use the “bootstrap” approach to simulate the critical value as suggested by Horowitz (2001). For example, one could assume the errors were normal and

draw repeated samples of  $Y$  data using least squares estimates of  $\pi$  and  $\Omega$  instead of the true population values. The  $1-\alpha$  quantile of the simulated distribution of  $W$  could then be used instead of  $c_\alpha$ . The existing theory of the bootstrap does not seem to cover the weak-instrument model studied here and few simulation studies have been conducted. This remains a promising area for further research.

- Bound, Jaeger and Baker (1995), Staiger and Stock (1997) and Hahn and Hausman (2002) propose two-step testing procedures. One first tests the hypothesis that  $\pi = 0$ . If this hypothesis is rejected, one uses the Wald statistic to test  $H_0$ . However, there is no clear advice as how the applied researcher should proceed if the pre-test indicates that the usual asymptotics is unreliable since identification is weak.
- Anderson and Rubin (1949) proposed to reject  $H_0$  for large values of the statistic

$$(3) \quad AR = \frac{u_0' N_Z u_0 / k}{u_0' M_Z u_0 / (n - k)}$$

where  $u_0 \equiv y_1 - y_2 \beta_0$ . Under normality  $AR$  has an F-distribution when the null hypothesis is true. Hence this test is exactly similar and has no size problems. However, if  $k > 1$ , the test has no particular optimum properties. Indeed, if  $k$  is large, the power of the Anderson-Rubin test can be quite low.

- A number of other tests that sacrifice power to guarantee correct size have been proposed. Staiger and Stock (1997) suggest the use of Bonferroni's inequality

to obtain a conservative test. Wang and Zivot (1998) suggest replacing the asymptotic critical value by some larger, conservative value guaranteeing that the null rejection probability is no larger than the nominal size. Dufour and Jasiak (2001) suggest a test with correct size using split-sample techniques. Unfortunately, all these procedures waste power unnecessarily when identification turns out to be strong.

## 4 Classical Testing Theory

Most of the econometric literature on testing in structural models has employed a large-sample asymptotic approach. However, it turns out that one can obtain some interesting results using the traditional exact theory as developed in Chapter 4 of the classical text by Lehmann (1986). This theory is based on the concept of a complete sufficient statistic.

**Definition 1** *Let  $\{P_\theta; \theta \in \Theta\}$  be a parametric model. A statistic  $T$  is sufficient for the family if the distribution of the data given  $T = t$  does not depend on the value for  $\theta$ .*

**Definition 2** *Let  $\mathcal{P}$  be a family of probability distributions for a statistic  $T$ . Then the statistic is said to be complete (with respect to this family) if  $E[f(T)] = 0$  for all  $P \in \mathcal{P}$  implies  $f(x) = 0$  almost surely.*

Finding tests with correct size is not trivial in multiparameter models when nuisance parameters are present under the null hypothesis. The task can be simplified if we find tests that have the Neyman  $\alpha$ -structure:

**Definition 3** *Let  $\{P_\theta; \theta \in \Theta\}$  be a parametric model and  $T$  a sufficient statistic in the submodel  $\{P_\theta; \theta \in \Theta_0\}$ . A test is said to have a Neyman  $\alpha$ -structure relative to  $T$  and  $\Theta_0$  if its rejection probability conditioned on  $T = t$  remains equal to  $\alpha$  for all values  $\theta$  in  $\Theta_0$ .*

Clearly, a test with Neyman structure will necessarily be similar. Moreover, if the submodel represents the distribution *under the null hypothesis* and if the statistic  $T$  is complete in that submodel, then every  $\alpha$ -similar test can also be represented in terms of Neyman  $\alpha$ -structure:

**Lemma 1** *If a statistic  $T$  is sufficient and complete in the submodel  $\{P_\theta; \theta \in \Theta_0\}$  then every  $\alpha$ -similar test of the hypothesis  $\theta \in \Theta_0$  also has the Neyman  $\alpha$ -structure with respect to  $T$  and  $\Theta_0$ .*

Since a test with the Neyman  $\alpha$ -structure with respect to a statistic  $T$  has constant conditional null rejection probability  $\alpha$  on each of the surfaces  $T = t$ , this result can essentially reduce the problem of finding similar tests to that of testing a simple hypothesis for each value of  $t$ .

In Chapters 2 and 3, this Lemma is used to find tests with correct size in our simple two-equation model. If the reduced-form errors are normal with *known* variance

matrix, the model is a member of the exponential family under the null hypothesis. We find a complete sufficient statistic  $T$  that satisfies the assumptions of the Lemma. This allow us to establish the *Similarity Condition* in Chapter 2 which characterizes all similar tests in terms of tests with the Neyman  $\alpha$ -structure. Moreover, it also allow us in Chapter 3 to find similar tests based on the conditional distribution of arbitrary nonpivotal statistics.

When the reduced-form covariance matrix  $\Omega$  is not known, it seems no longer possible to find a complete sufficient statistic that satisfies the assumptions of the Lemma. However, as noted by Anderson, Kunitomo and Sawa (1982) in a different context, although the simultaneous equations model with known covariance matrix has a much simpler mathematical structure than the model with unknown covariance matrix, inference procedures for the two models behave very much alike in moderate sized samples. As shown in Chapter 4, replacing the unknown  $\Omega$  by an estimate has little effect on the size and power of the tests. The exact theory developed in Chapters 2 and 3 suggests a way to find tests that are approximately similar in the presence of weak instruments even when  $\Omega$  is unknown.

## Chapter 2

# Tests with Correct Size when Instruments May Be Arbitrarily Weak

This chapter develops a methodology for finding similar tests for the structural parameter under the assumption of normal disturbances with *known* covariance matrix  $\Omega$ . A power bound for similar tests is derived and it is shown that there exists no optimal test when the model is overidentified. A family of similar tests based on pivotal statistics is examined. This family includes a particular score test independently analyzed by Kleibergen (2001). Although this test is asymptotically optimal under local alternatives, it appears to have poor power for some parts of the parameter space.

# 1 The Family of Similar Tests

Ignoring an additive constant, the log-likelihood function can be written as

$$l(Y; \beta, \pi, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} \text{tr} (\Omega^{-1} V' V)$$

where  $V = Y - Z\pi(\beta, 1)$ . Using the factorization theorem, we note that  $Z'Y$  is a sufficient statistic for the unknown parameters  $\beta$  and  $\pi$  when  $\Omega$  is known. For any non-singular  $2 \times 2$  matrix  $D$ , the two columns of  $Z'YD$  are also a pair of sufficient statistics. A convenient choice is the pair

$$(1) \quad S = Z'Yb \quad \text{and} \quad T = Z'Y\Omega^{-1}a$$

where  $a = (\beta_0, 1)'$ . Although the null distribution of  $S = Z'u_0$  does not depend on the nuisance parameter  $\pi$ , the null distribution of  $T$  is very sensitive to the value of  $\pi$ . Indeed, a little algebra shows that  $T$  is closely related to  $\hat{\pi}$ , the maximum likelihood estimate of  $\pi$  when  $\beta$  is constrained to take the null value  $\beta_0$ :

$$T = a'\Omega^{-1}a \cdot Z'Z\hat{\pi}.$$

Indeed,  $T$  is a sufficient statistic for  $\pi$  under the null hypothesis  $H_0 : \beta = \beta_0$ . The vectors  $S$  and  $T$  are independent and normally distributed under both the null and

alternative hypotheses. Specifically,

$$S \sim N(Z'Z\pi(\beta - \beta_0), Z'Z \cdot b'\Omega b)$$

$$T \sim N(Z'Z\pi \cdot \bar{a}'\Omega^{-1}a, Z'Z \cdot a'\Omega^{-1}a)$$

where  $\bar{a} = (\beta, 1)$ .

Because they are sufficient statistics, all tests can be written as (possibly randomized) functions of  $S$  and  $T$ . Specifically, let  $\phi$  be a critical function such that for each  $S$  and  $T$  the test rejects or accepts the null with probabilities  $\phi(S, T)$  and  $1 - \phi(S, T)$ , respectively. Let  $E_0$  represent expectation over the distribution of  $S$  under the null  $H_0 : \beta = \beta_0$  and suppose that the set  $P$  in which  $\pi$  is known to lie contains a  $k$ -dimensional rectangle so that the statistic  $T$  is complete under the null hypothesis. Since  $T$  is also sufficient under the null hypothesis, we can apply Lemma 1 of Chapter 1 to show that every  $\alpha$ -similar test has the Neyman  $\alpha$ -structure in our model:

**Theorem 1 (Similarity Condition)** *A test is similar at size  $\alpha$  if and only if it can be written as  $\phi(S, T)$  such that  $E_0\phi(S, t) = \alpha$  for almost every  $t$ .*

**Proof.** Since randomization is allowed, any test can be written as  $\phi(S, T)$ . Since

the test is similar at size  $\alpha$ , it must be the case that:

$$(2) \quad E_0 \phi(S, T) = \alpha \quad , \forall \pi \in P$$

By Lemma A.1 in the Appendix, the family of distributions of  $T$  when the null hypothesis is true,  $\mathcal{P}^T = \{P_{\beta_0, \pi}^T; \pi \in P\}$ , is complete. Consequently, the following holds:

$$E_0 \{\phi(S, T) | t\} = \alpha \quad , a.e. \mathcal{P}^T$$

Note that the distribution of  $S$  does not depend on  $\pi$  under the null hypothesis and that  $S$  is independent of  $T$ . Therefore, using also the fact that  $\phi$  is integrable:

$$(3) \quad E_0 \phi(S, t) = \alpha \quad , a.e. \mathcal{P}^T$$

Conversely, if the test is such that (3) holds, then (2) is trivially true. Therefore, the test is similar at size  $\alpha$ . ■

## 1.1 Similar Tests Based on Pivotal Statistics

A trivial example that satisfies the *Similarity Condition* is the test that rejects  $H_0$  for large values of the Anderson-Rubin statistic (modified to take into account  $\Omega$  is known)

$$AR_0 = S' (Z'Z)^{-1} S / \sigma_0^2$$

where  $\sigma_0^2 = b'\Omega b$  is the variance of each element of  $u_0 \equiv y_1 - y_2\beta_0$ . But there are also similar tests that use the additional information contained in  $T$ . Consider, for example, the family of pivotal test statistics

$$(4) \quad \mathcal{T}_g = \frac{g(T)'S}{\sigma_0 \sqrt{g(T)'Z'Zg(T)}}$$

where  $g$  is a measurable mapping from  $\mathbf{R}^k$  onto  $\mathbf{R}^k$ . These statistics are  $N(0, 1)$  under  $H_0$ . For the one-sided alternative  $\beta > \beta_0$ , a similar test at significance level  $\alpha$  rejects  $H_0$  if  $\mathcal{T}_g > z_\alpha$ , the  $(1 - \alpha)$  standard normal quantile. For a two sided-hypothesis, a similar test at significance level  $\alpha$  rejects  $H_0$  if  $|\mathcal{T}_g| > z_{\alpha/2}$ .

**Example 1** Let  $g(T) = (Z'Z)^{-1}T/a'\Omega^{-1}a = \hat{\pi}$  so

$$\mathcal{T}_g = \frac{\hat{\pi}'S}{\sigma_0 \sqrt{\hat{\pi}'Z'Z\hat{\pi}}}$$

*The power properties of this test and its extensions when  $\Omega$  is unknown will be studied in Chapters 3 and 4.*

**Example 2** Let  $g(x) \equiv d$  for any  $x \in \mathbf{R}^k$ . In this case,  $\mathcal{T}_g$  is given by

$$\mathcal{T}_g = \frac{d'S}{\sigma_0 \sqrt{d'Z'Zd}}$$

Later, it will be shown that tests based on this pivotal statistical are optimal when  $d$  happens to be a positive multiple of  $\pi$ . This test can be used in practice if there is strong prior belief about the coefficients on the instruments.

**Example 3** Let the  $j$ th element of  $g$  be one if  $|\hat{\pi}_j|$  is the largest among  $|\hat{\pi}_1|, \dots, |\hat{\pi}_k|$  and zero otherwise. Tests based on this statistic should have good power properties when only one instrument is valid, but it is not known which one.

Not all pivotal statistics are members of the  $\mathcal{T}_g$  family. For example, Staiger and Stock (1997) suggest the possibility of splitting the sample into two half samples to construct similar tests, each sub-sample consisting of  $\{y_1^{(j)}, y_2^{(j)}, Z^{(j)}\}$  for  $j = 1, 2$ . Consider the statistic

$$\mathcal{J}_{12} = \frac{\hat{\pi}_{ols}^{(1)'} S^{(2)}}{\sigma_0 \sqrt{\hat{\pi}_{ols}^{(1)'} Z^{(2)'} Z^{(2)} \hat{\pi}_{ols}^{(1)}}}$$

where  $\hat{\pi}_{ols}^{(1)} \equiv \left(Z^{(1)'} Z^{(1)}\right)^{-1} Z^{(1)'} y_2^{(1)}$  is the *OLS* estimator of  $\pi$  using the first sub-sample and  $S^{(2)} \equiv Z^{(2)'}(y_1^{(2)} - y_2^{(2)}\beta_0)$ . An important feature is that  $\hat{\pi}_{ols}^{(1)}$  is independent of  $S^{(2)}$ , as long as the observations from the two sub-samples are independent. Therefore,  $\mathcal{J}_{12}$  is pivotal and tests based on  $\mathcal{J}_{12}$  are similar at level  $\alpha$  if the appropriate normal critical value is used. However, this test wastes power since  $\mathcal{J}_{12}$  requires randomization to be expressed as a function of  $S$  and  $T$ .

## 1.2 Pre-testing Procedures

Example 3 can be interpreted as a pre-test procedure: one first decides on the basis of  $\hat{\pi}$  which instrument is best and then tests  $H_0$  using a one-instrument version of the Anderson-Rubin test. Although pre-test procedures are commonly used in econometrics, the fact that the first-step typically affects the size of the second-step test is usually ignored. Pre-testing on the basis of  $\hat{\pi}$  however does not cause any difficulties with tests based on the pivotal statistics  $\mathcal{T}_g$ . More generally, we have the following implication of Theorem 1:

**Proposition 1** *Let  $h(T)$  be a measurable real valued function and let  $\phi_1(S, T)$  and  $\phi_2(S, T)$  be two similar tests at level  $\alpha$ . Finally, let  $\phi_3 = I[h(T) > c] \phi_1 + I[h(T) \leq c] \phi_2$  where  $I$  is the indicator function taking the value one if the argument is true and zero otherwise. Then  $\phi_3$  is also a similar test at level  $\alpha$ .*

For example, one might decide to use the Anderson-Rubin test if  $\hat{\pi}$  is near the origin and use the test of Example 1 if  $\hat{\pi}$  is far from the origin. If the decision is based on the reduced-form “ $F$ -statistic”  $a' \Omega^{-1} a \cdot \hat{\pi}' Z' Z \hat{\pi}$ , the procedure is valid. That is, choosing which similar test to be used after testing if  $\pi$  is significantly different from zero does not affect the final test’s size as long as the preliminary test is based on the *constrained* maximum likelihood estimate.

## 2 Power Functions

When  $\pi$  is far from the origin and the sample size is large, the standard likelihood ratio, Wald, and Lagrange multiplier two-sided tests of the hypothesis  $\beta = \beta_0$  are approximately best unbiased and have approximate power

$$(5) \quad 1 - G \left( c_\alpha ; \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\sigma_0^2} \right)$$

where  $c_\alpha$  is the  $1 - \alpha$  quantile of a central  $\chi^2(1)$  distribution and  $G(\cdot; \mu)$  is the noncentral  $\chi^2(1)$  distribution function with noncentrality parameter  $\mu$ . However, these tests are not generally similar and the power approximation is unreliable when  $\pi$  is near the origin. Only in the case  $k = 1$  where the model is exactly identified, do we have an exact optimality result. Then, as shown in the Appendix, the Anderson-Rubin  $AR_0$  test is uniformly most powerful unbiased and has exact power function given by (5). Therefore, it is not surprising why Monte Carlo simulations ran by Wang and Zivot (1998) and Zivot, Startz and Nelson (1998) suggests that no test dominates the one proposed by Anderson and Rubin (1949) when  $k = 1$ . Its power is very close to the power of the Anderson-Rubin  $AR_0$  test, which is itself the optimal test when  $\Omega$  is known.

When  $k > 1$ , there exists no uniformly most powerful test. To assess the power properties of the similar tests described earlier, it is useful to find the *power envelope*, the upper bound for the rejection probability for each alternative  $\beta \neq \beta_0$  and value of

$\pi$ . Moreover, one could find the power upper bound *within* the class of similar tests.

In the appendix we show:

**Theorem 2** *For testing  $H_0 : \beta = \beta_0$  against  $H_1 : \beta \neq \beta_0$  when  $\Omega$  is known and*

*$\pi \neq 0$ , we have*

a. *If the model is just identified, the uniformly most power unbiased test has a power function given by*

$$(6) \quad P_{\beta,\pi}(AR_0 > c_\alpha) = 1 - G \left( c_\alpha ; \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\sigma_0^2} \right)$$

b. *If  $\pi$  is known, the uniformly most powerful unbiased test has a power function is given by*

$$(7) \quad P_{\beta,\pi}(\mathcal{R} > c_\alpha) = 1 - G \left( c_\alpha ; \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\omega_{11} - \omega_{12}^2 / \omega_{22}} \right)$$

where  $R$  is defined in equation A.3.

c. *If  $\pi$  unknown and  $P$  contains a  $k$ -dimensional rectangle, the two-sided power envelope for the class of exactly similar tests is given by*

$$(8) \quad P_{\beta,\pi} \left( \frac{(\pi' S)^2}{\sigma_0^2 \pi' Z' Z \pi} > c_\alpha \right) = 1 - G \left( c_\alpha ; \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\sigma_0^2} \right)$$

**Proof.** See the Appendix. ■

Note that (7) is an upper bound for the power of any two-sided test with correct size. Since  $\sigma_0^2$  is necessarily smaller than  $\omega_{11} - \omega_{12}^2/\omega_{22}$  when  $u$  is correlated with  $v_2$ , insisting on similarity lowers the attainable power of the test. The optimal test for known  $\pi$  can be understood as the optimal similar test when the nuisance-parameter set contains only one element; the loss in power is then due to increase in the nuisance parameter space.

### 3 Score Test

Theorem 2 suggests that replacing  $\pi$  in (8) by an estimate might lead to a reasonable test. Using the OLS estimator  $(Z'Z)^{-1}Z'y_2$  does not produce a similar test. But, as already suggested in our Example 1, using the constrained maximum likelihood estimator does. It is shown in the Appendix that the gradient of the log likelihood function with respect to  $\beta$ , when evaluated at  $(\beta_0, \hat{\pi})$ , is proportional to  $\hat{\pi}'S$ . Hence, we have the following result:

**Theorem 3** *The test that rejects the null if*

$$(9) \quad LM_0 = \frac{(\hat{\pi}'S)^2}{\sigma_0^2 \hat{\pi}'Z'Z\hat{\pi}}$$

*is larger than  $c_\alpha$  is a Lagrange multiplier (or score) test based on the normal likelihood*

with  $\Omega$  known.

**Proof.** The derivative of the log likelihood function with respect to  $\beta$  evaluated at  $\beta_0$  and at  $\hat{\pi}$  is given by

$$\frac{\omega_{12}\hat{\pi}'Z'(y_2 - Z\hat{\pi}) - \omega_{22}\hat{\pi}'Z'(y_1 - Z\hat{\pi}\beta_0)}{\omega_{11}\omega_{22} - \omega_{12}^2}$$

which can be rewritten as

$$-\frac{\hat{\pi}'Z'(y_1 - y_2\beta_0)}{\omega_{11} + \omega_{22}\beta_0^2 - 2\omega_{12}\beta_0}.$$

■

Although  $\hat{\pi}$  is an unbiased estimator of  $\pi$  if the null hypothesis is true, it is biased under the alternatives  $\beta \neq \beta_0$ . In fact,  $E(\hat{\pi}) = \pi d$ , where the scalar  $d$  is given by

$$(10) \quad d = \frac{\omega_{11} - \omega_{12}(\beta + \beta_0) + \omega_{22}\beta\beta_0}{\omega_{11} - 2\omega_{12}\beta_0 + \omega_{12}\beta_0^2}.$$

The fact that  $\hat{\pi}$  is a biased estimator of  $\pi$  under the alternative hypothesis does not necessarily imply bad power properties for the  $LM_0$  test. The  $LM_0$  test fails to have good power only when the direction of  $\pi$ , thought as a vector, is not estimated accurately. In the extreme case  $d = 0$ ,  $(Z'Z)^{-1/2}\hat{\pi}$  will randomly pick equally likely directions, regardless of the true value of the nuisance parameter  $\pi$ . This suggests

that the test will have poor power properties whenever  $d$  is near zero.

## 4 Conclusions

In this Chapter, we characterized the class of similar tests and derived its power upper bound. We also proposed a class of similar tests based on pivotal statistics. A member of this class is a particular score test. This test is (locally) asymptotically optimal but we indicate that this test has poor power for some parts of the parameter space. Improved inference in the weak instrument case might be possible by exploring the properties of other tests that satisfy the *Similarity Condition* of Theorem 1 or by using the pre-testing procedure of Proposition 1.

## Chapter 3

# A Conditional Likelihood Ratio

## Test for Structural Models

This Chapter develops a general procedure for constructing valid tests of structural coefficients based on the conditional distribution of nonpivotal statistics. The conditional approach is employed to find critical value functions for Wald and likelihood ratio tests yielding correct rejection probabilities no matter how weak the instruments.

Although the conditional Wald, Anderson-Rubin and score tests have relatively poor power in some regions of the parameter space, the conditional likelihood ratio test has overall good power properties. Monte Carlo simulations suggest that this conditional likelihood ratio test not only has power close to the power envelope of similar tests when identification is good but it also dominates the test proposed by

Anderson and Rubin (1949) and the tests proposed in Chapter 2 when identification is weak.

## 1 Similar Tests Based on Nonpivotal Statistics

Any test statistic that depends only on  $S$  will be pivotal and can be used to form a similar test. Likewise, as discussed in Chapter 2, similar tests can be constructed from pivotal statistics of the form  $g(T)'S/\sqrt{g(T)'Z'Zg(T)}$  where  $g$  is any (measurable)  $k$ -dimensional vector depending on  $T$ . The goal here instead is to find a similar test at level  $\alpha$  based on a nonpivotal test statistic  $\psi(S, T, \Omega, \beta_0)$ . The following approach is suggested by the analysis in Lehmann (1986, Chapter 4). Although the marginal distribution of  $\psi$  depends on  $\pi$ , the conditional null distribution of  $\psi$  given that  $T$  takes on the value  $t$  does not depend on  $\pi$ . As long as this distribution is continuous, its quantiles can be computed and used to construct a similar test. The following proposition follows immediately:

**Proposition 1** *Suppose that  $\psi(S, t, \Omega, \beta_0)$  is a continuous random variable under  $H_0$  for every  $t$ . Define  $c(t, \Omega, \beta_0, \alpha)$  to be the  $1 - \alpha$  quantile of the null distribution of  $\psi(S, t, \Omega, \beta_0)$ . Then the test that rejects  $H_0$  if  $\psi(S, T, \Omega, \beta_0) > c(T, \Omega, \beta_0, \alpha)$  is similar at level  $\alpha \in (0, 1)$ .*

**Proof.** In fact we need only assume that  $\psi(S, t, \Omega, \beta_0)$  is a continuous random variable for all  $t$  except for a set having  $T$ -probability zero. For any  $t$  where

$\psi(S, t, \Omega, \beta_0)$  is not a continuous random variable, define  $c_\alpha(t)$  to be zero. Otherwise, let  $c_\alpha(t)$  be the  $1 - \alpha$  quantile of  $\psi$ . Thus,  $\Pr[\psi(S, T, \Omega, \beta_0) > c_\alpha(T) | T = t] = \alpha$  by construction. Since this holds for almost all  $t$ ,  $\Pr[\psi(S, T, \Omega, \beta_0) > c_\alpha(T)] = \alpha$  unconditionally. ■

It is shown in Chapter 2 that  $c_\psi(T, \Omega, \beta_0, \alpha)$  does not depend on  $T$  when  $\psi$  is pivotal. Thus, the conditional approach for finding a similar test may be thought of as replacing the nonpivotal statistic  $\psi(S, T, \Omega, \beta_0)$  by the new statistic  $\psi(S, T, \Omega, \beta_0) - c_\psi(T, \Omega, \beta_0, \alpha)$ . Alternatively, since conditioning on  $T$  is the same as conditioning on  $\hat{\pi}$ , this approach may be interpreted as adjusting the critical value based on a preliminary estimate of  $\pi$ . Henceforth,  $c_\psi(T, \Omega, \beta_0, \alpha)$  will be referred to as the critical value function for the test statistic  $\psi$ .

## 1.1 Examples of Conditional Tests

To illustrate the conditional approach, we now consider a number of examples.

**Example 1** *The Anderson-Rubin statistic for known  $\Omega$  is*

$$AR_0 = S'(Z'Z)^{-1}S/\sigma_0^2.$$

*This statistic is distributed chi-square- $k$  under the null hypothesis and it is consequently pivotal. Its conditional distribution given  $T = t$  does not depend on  $t$  and its*

*critical value function collapses to a constant*

$$c_{AR}(t, \Omega, \beta_0, \alpha) = q_\alpha(k)$$

*where  $q_\alpha(df)$  is the  $1 - \alpha$  quantile of a chi-square distribution with  $df$  degrees of freedom. As argued earlier, although the Anderson-Rubin test has correct size, it has poor power properties when the model is over-identified since the degrees of freedom is larger than the number of parameters being tested.*

**Example 2** *The score statistic in Chapter 2 is given by:*

$$LM_0 = \frac{(S'\hat{\pi})^2}{\sigma_0^2\hat{\pi}'Z'Z\hat{\pi}}.$$

*As mentioned before,  $\hat{\pi}$  is independent of  $S$ . Thus, the null distribution of  $LM_0$  is chi-square-1 and a test that rejects  $H_0$  for large values of  $LM_0$  has correct size as long as the appropriate critical value is used. Again, the score test statistic is pivotal and its critical value function collapses to a constant*

$$c_{LM}(t, \Omega, \beta_0, \alpha) = q_\alpha(1) .$$

*Like the Wald and likelihood ratio tests, the score test is (locally) asymptotically optimal when the structural parameters are identified.*

**Example 3** *The Wald statistic centered around the 2SLS estimator is given by*

$$W_0 = (b_{2SLS} - \beta_0)' Y_2' N_Z Y_2 (b_{2SLS} - \beta_0) / \hat{\sigma}^2$$

where  $b_{2SLS} = (y_2' N_Z y_2)^{-1} y_2' N_Z y_1$  and  $\hat{\sigma}^2 = [1 - b_{2SLS}] \Omega [1 - b_{2SLS}]'$ . Here, the non-standard structural error variance estimate exploits the fact that  $\Omega$  is known. The critical value function for  $W_0$  can be written as

$$c_W(T, \Omega, \beta_0, \alpha) = \bar{c}_W(\tau, \Omega, \beta_0, \alpha)$$

where  $\tau \equiv (A_0' \Omega^{-1} A_0)^{-1/2} t'(Z'Z)^{-1} t(A_0' \Omega^{-1} A_0)^{-1/2}$ .

**Example 4** *The likelihood ratio statistic, for known  $\Omega$ , is defined as*

$$LR_0 = 2 \left[ \max_{\beta, \Pi} l(Y; \beta, \Pi, \Omega) - \max_{\Pi} l(Y; \beta_0, \Pi, \Omega) \right]$$

where  $l$  is the log likelihood function after concentrating out  $\delta$  and  $\Gamma$ . Various authors have noticed that, in curved exponential models, the likelihood ratio test performs well for a wide range of alternatives; see, for example, Van Garderen (2000). In the appendix we show that the critical value function for the likelihood ratio test has the form

$$c_{LR}(T, \Omega, \beta_0, \alpha) = \bar{c}_{LR}(\tau, \alpha);$$

that is, it is independent of  $\Omega$  and  $\beta_0$ .

To implement the conditional test based on a nonpivotal statistic  $\psi$ , we need to be able to compute the conditional quantile  $c_\psi(t, \Omega, \beta_0, \alpha)$ . Although in principle the entire critical value function can be derived from the known null distribution of  $S$ , for most choices for  $\psi$  a simple analytical expression seems out of reach. A Monte Carlo simulation from the null distribution of  $S$  is much simpler. Indeed, the applied researcher need only do a simulation for the actual value  $t$  observed in the sample and for the particular  $\beta_0$  being tested; there is no need to derive the whole critical value function  $c_\psi(t, \Omega, \beta_0, \alpha)$ .

The critical value function for a given test statistic  $\psi$  will generally depend on the  $k$ -dimensional vector  $t$ . However, as noted above, the critical value functions for the Wald and likelihood ratio statistics depend on  $t$  only through the scalar  $\tau$ . This can be a considerable simplification when  $k$  is large.

## 2 The Conditional Likelihood Ratio Test

In this Section, we work out in more detail expressions for the conditional likelihood ratio test. When  $\Omega$  is known, the likelihood ratio statistic, defined to be two times the log of the likelihood ratio, is given by:

$$(1) \quad LR_0 = \bar{\lambda}^{\max} - \frac{a'\Omega^{-1}Y'N_ZY\Omega^{-1}a}{a'\Omega^{-1}a}$$

where  $\bar{\lambda}^{\max}$  is the largest eigenvalue of  $\Omega^{-1/2}Y'N_ZY\Omega^{-1/2}$ . This statistic can be written as a function of the sufficient statistics  $S$  and  $T$ . However, the expression is somewhat simpler when written in terms of the standardized statistics

$$\bar{S} = \frac{(Z'Z)^{-1/2} S}{\sqrt{b'\Omega b}} \quad \text{and} \quad \bar{T} = \frac{(Z'Z)^{-1/2} T}{\sqrt{a'\Omega^{-1}a}}$$

having covariance matrices equal to the identity matrix. Under the null hypothesis,  $\bar{S}$  has mean zero so  $\bar{S}'\bar{S}$  is distributed as chi square with  $k$  degrees of freedom. The statistic  $\bar{T}'\bar{T}$  is distributed as noncentral chi square with noncentrality proportional to  $\pi'Z'Z\pi$ ; it can be viewed as a natural statistic for testing the hypothesis that  $\pi = 0$  under the assumption that  $\beta = \beta_0$ .

In the Appendix, the following expression for the likelihood ratio statistic is derived:

$$(2) \quad LR_0 = \frac{1}{2} \left[ \bar{S}'\bar{S} - \bar{T}'\bar{T} + \sqrt{[\bar{S}'\bar{S} + \bar{T}'\bar{T}]^2 - 4[\bar{S}'\bar{S} \cdot \bar{T}'\bar{T} - (\bar{S}'\bar{T})^2]} \right].$$

When  $k = 1$ ,  $\bar{S}$  and  $\bar{T}$  are scalars and the  $LR_0$  statistic collapses to the pivotal statistic  $\bar{S}'\bar{S}$ . In the overidentified case the  $LR_0$  statistic depends also on  $\bar{T}'\bar{T}$  and is no longer pivotal. Nevertheless, a similar test can be found by applying Proposition 1. Again, an analytic expression for the critical value function for the  $LR_0$  test statistic is not available but the needed values can be computed by simulation. Some general properties of the function are known.

**Proposition 2** *The critical value function for the conditional  $LR_0$  test depends only on  $\alpha$ ,  $k$ , and  $\bar{t}'\bar{t}$ . It satisfies*

$$\bar{c}_{LR}(\bar{t}'\bar{t}, k, \alpha) \rightarrow q_\alpha(1) \text{ as } \bar{t}'\bar{t} \rightarrow \infty$$

$$\bar{c}_{LR}(\bar{t}'\bar{t}, k, \alpha) \rightarrow q_\alpha(k) \text{ as } \bar{t}'\bar{t} \rightarrow 0$$

**Proof.** Note that  $(\bar{S}, \bar{T}) = (Z'Z)^{-1/2}Z'Y\Omega^{-1/2}W$  where

$$W = \begin{bmatrix} \frac{\Omega^{1/2}b}{\sqrt{b'\Omega b}}, \frac{\Omega^{-1/2}a}{\sqrt{a'\Omega^{-1}a}} \end{bmatrix}$$

is an orthogonal matrix. Thus the eigenvalues of  $\Omega^{-1/2}Y'N_ZY\Omega^{-1/2}$  are the same as the eigenvalues of  $(\bar{S}, \bar{T})'(\bar{S}, \bar{T})$ . The largest eigenvalue then is given by

$$\bar{\lambda}^{\max} = \frac{1}{2} \left[ \bar{T}'\bar{T} + \bar{S}'\bar{S} + \sqrt{(\bar{T}'\bar{T} + \bar{S}'\bar{S})^2 - 4[\bar{S}'\bar{S} \cdot \bar{T}'\bar{T} - (\bar{S}'\bar{T})^2]} \right]$$

Moreover,  $a'\Omega^{-1}Y'NY\Omega^{-1}a/a'\Omega^{-1}a = \bar{T}'\bar{T}$ . Therefore, the  $LR_0$  test statistic is given by expression (5). For  $\bar{T}'\bar{T} \neq 0$ ,  $LR_0$  can be rewritten as

$$LR_0 = \frac{1}{2} \left[ Q_1 + Q_{k-1} - \bar{T}'\bar{T} + \sqrt{(Q_1 + Q_{k-1} + \bar{T}'\bar{T})^2 - 4Q_{k-1} \cdot \bar{T}'\bar{T}} \right]$$

where  $Q_1 = \bar{S}'\bar{T}(\bar{T}'\bar{T})^{-1}\bar{T}'\bar{S}$  and  $Q_{k-1} = \bar{S}'[I - \bar{T}(\bar{T}'\bar{T})^{-1}\bar{T}']\bar{S}$ . Conditioned on  $\bar{T} = \bar{t}$ ,  $Q_1$  and  $Q_{k-1}$  are independent and under  $H_0$  have chi-square distributions with one

and  $k - 1$  degrees of freedom, respectively. Therefore the critical value function just depends on  $k$ ,  $\alpha$  and  $\bar{\ell}'\bar{\ell}$ . When  $\bar{T} = 0$ ,  $LR_0 = \bar{S}'\bar{S}$ , which is a chi-square- $k$  random variable. When  $\bar{T}'\bar{T} \rightarrow \infty$ ,  $LR_0 \rightarrow (\bar{S}'\bar{T})^2/\bar{T}'\bar{T}$ , which is a chi-square-one random variable. ■

Table 1 presents the critical value function calculated from 10,000 Monte Carlo replications for the significance level of 5%. When  $k = 1$ , the true critical value function is a constant equal to 3.84 at level 5%. The slight variation in the first column of Table 1 is due to simulation error. For each  $k$ , the critical value function has approximately an exponential shape, decreasing from  $q_\alpha(k)$  at  $\bar{\ell}'\bar{\ell} = 0$  to  $q_\alpha(1)$  as  $\bar{\ell}'\bar{\ell}$  tends to infinity. For example, when  $k = 4$ , the approximation  $\bar{c}(\bar{\ell}'\bar{\ell}, 4, 0.05) = 3.84 + 5.65 \exp(-\bar{\ell}'\bar{\ell}/7)$  seems to fit reasonably well.

The shape of the critical value function explains why the method proposed by Wang and Zivot (1998) leads to a test with low power. Their proposed critical value of a chi-square- $k$  quantile is the upper bound for the true critical value function. The method proposed here can also be seen as a refinement of the method proposed by Zivot, Startz and Nelson (1998) that selects for the critical value either  $q_\alpha(k)$  or  $q_\alpha(1)$  depending on a preliminary test of the hypothesis  $\pi = 0$ . The conditional approach has the advantage that it is not *ad hoc* and the final test has correct size without unnecessarily wasting power. Figure 1 illustrates each method, sketching its respective critical values<sup>1</sup> for different values of  $\bar{\ell}'\bar{\ell}$  when the number of instruments

---

<sup>1</sup>The pre-testing procedure proposed by Zivot, Nelson and Startz (1998) is based on the *OLS* estimator for  $\pi$ . Instead, figure 1 sketches the critical value function based on a pre-testing on the

equals four.

### 3 Monte Carlo Simulations

To evaluate the power of the conditional  $W_0$  and  $LR_0$  tests, a 1,000 replication experiment was performed based on design I of Staiger and Stock (1997). The hypothesized value  $\beta_0$  is zero. The elements of the  $100 \times 4$  matrix  $Z$  are drawn as independent standard normal and then held fixed over the replications. Two different values of the  $\pi$  vector are used so that the “population” first-stage F-statistic (in the notation of Staiger and Stock)  $\lambda'\lambda/k = \pi'Z'Z\pi/(\omega_{22}k)$  takes the values 1 (weak instruments) and 10 (good instruments). The rows of  $(u, v_2)$  are i.i.d. normal random vectors with unit variances and correlation  $\rho$ . Results are reported for  $\rho$  taking the values 0.00, 0.50 and 0.99. The critical values for the conditional likelihood ratio and Wald tests were based on 1,000 replications.

In addition to the two conditional tests, denoted as  $LR_0^*$  and  $W_0^*$ , two other similar tests were evaluated: the Anderson-Rubin test based on the statistic  $AR_0 = \bar{S}'\bar{S}$  (modified to take into account that  $\Omega$  is known) and the score test based on the statistic  $LM_0 = (\bar{T}'\bar{S})^2/\bar{T}\bar{T}$ . Figures 2 and 3 graph, for a fixed value of  $\pi$  and  $\rho$ , the rejection probabilities of the  $AR_0$ ,  $LM_0$ , conditional  $LR_0$  and conditional  $W_0$  tests as functions of the true value  $\beta$ .<sup>2</sup> The power envelope for similar tests is also included.

---

constrained maximum likelihood estimator for  $\pi$ .

<sup>2</sup>As  $\beta$  varies,  $\omega_{11}$  and  $\omega_{12}$  change to keep the structural error variance and the correlation between  $u$  and  $v_2$  constant.

In each figure, each power curve is at approximately the 5% level when  $\beta$  equals  $\beta_0$ . This reflects the fact that each test is similar. As expected, the power curves become steeper as the quality of instruments improve.

Both figures show that the  $AR_0$  test has power considerably lower than the envelope power. The  $LM_0$  test has relative low power either for the weak-instrument case or for some alternatives  $\beta$  for the good-instrument case. The  $W_0$  test is biased, reflecting the finite-sample bias of the  $2SLS$  estimator. These poor power properties are not shared by the conditional  $LR_0$  test. The conditional likelihood ratio test not only seems to dominate the Anderson-Rubin test and the score test, but it also has power essentially equal to the power envelope for similar tests<sup>3</sup> when identification is strong.

## 4 Conclusions

Monte Carlo simulations suggest that the conditional likelihood ratio test has good size and power properties. If identification is good, this test has power curve essentially equal to the upper bound power curve for similar tests. Monte Carlo simulations also suggest that this test dominates the test proposed by Anderson and Rubin (1949) and other tests studied in Chapter 2.

The conditional approach used here for finding similar tests based on nonpivotal

---

<sup>3</sup>Other tests that have been proposed in the literature such as the Wald test based on the LIML estimator and the  $GMM_0$  test proposed by Wang and Zivot (1998) were also considered. However, Monte Carlo simulations suggest that their conditional counterparts have power no larger than the conditional likelihood ratio test.

statistics can be applied to other statistical problems involving nuisance parameters. Improved inference should be possible whenever a subset of the statistics employed to form a test statistic has a nuisance-parameter free distribution and is independent of the remaining statistics under the null hypothesis.

## Chapter 4

# Approximations when the Error

# Distribution is Unknown

In this Chapter, we extend the conditional approach for the case the error distribution is unknown by substituting a consistent estimate for  $\Omega$ . Monte Carlo evidence and weak-instrument asymptotics show that this extension does not affect significantly the size and power of the resulting test. In particular, we are able to address the *Dufour critique* by constructing informative confidence regions with approximately correct nominal level. In addition, we show how (approximately) similar tests can be obtained when the structural equation contains more than one explanatory variable.

# 1 Conditional Approach Revisited

In practice, of course, the reduced-form covariance matrix is unknown. Furthermore, there is no compelling reason to believe that the errors are exactly normally distributed. However, since  $\Omega$  can be well estimated even when identification is weak and standardized sums of independent random variables should be approximately normal, it seems plausible that, after replacing  $\Omega$  by a consistent estimate, the tests developed in Chapter 3 will behave well in moderate sized samples even with nonnormal errors. To illustrate the conditional approach when the variance of the disturbances is unknown, we consider a few examples.

## 1.1 The Anderson-Rubin Test

Replacing unknown  $\Omega$  by the estimate  $\hat{\Omega} = Y'(I - Z(Z'Z)^{-1}Z'Y)/(n - k)$  in  $AR_0$  gives the test statistic proposed by Anderson and Rubin (1949):

$$(1) \quad AR = \frac{u_0' N_Z u_0 / k}{u_0' M_Z u_0 / (n - k)}$$

This particular test statistic is exactly pivotal when the disturbances are normal with unknown variance. More importantly, this test statistic is approximately pivotal even when the assumption of normality is dropped. Consequently, it is expected that the Anderson-Rubin test which rejects the null for large values of the  $AR$  test statistic has good size properties even for nonnormal error distributions.

## 1.2 The Score Test

There are two natural estimates to be used in the score test: the OLS estimate  $\widehat{\Omega} = Y'(I - Z(Z'Z)^{-1}Z'Y)/(n - k)$  and the estimate that maximizes the likelihood function when  $\beta$  is constrained to equal the hypothesized value  $\beta_0$ . In particular, the score test statistic  $LM_0$  can be modified as

$$(2) \quad LM_1 = \frac{(\tilde{\pi}'S)^2}{\tilde{\sigma}^2 \tilde{\pi}'Z'Z\tilde{\pi}}$$

where  $\tilde{\pi} = (Z'Z)^{-1}Z'Y\widehat{\Omega}^{-1}b$  and  $\tilde{\sigma}^2 = b'\widehat{\Omega}b$ ; or alternatively as

$$(3) \quad LM_2 = \frac{(\pi_0'S)^2}{n^{-1}u_0'u_0 \cdot \pi_0'Z'Z\pi_0}$$

where  $\pi_0 = (Z'M_0Z)^{-1}Z'M_0y_2$  is the constrained MLE for  $\pi$  and  $M_0 = I - u_0(u_0'u_0)^{-1}u_0'$ .

The  $LM_1$  test has been independently proposed by Kleibergen (2001).

The  $LM_1$  and  $LM_2$  tests can both be interpreted as score tests. Let  $V = (y_1 - Z\pi\beta, y_2 - Z\pi)$ . Then the term  $\tilde{\pi}'S$  appearing in the numerator of (2) is just the gradient with respect to  $\beta$  of the objective function

$$Q(\beta, \pi) = tr(\widehat{\Omega}^{-1}V'V)$$

evaluated at the constrained maximizing value  $(\beta_0, \tilde{\pi})$ . The term  $\pi_0'S$  appearing in the

numerator of (3) is just the gradient with respect to  $\beta$  of the log likelihood function

$$L(\beta_0, \pi, \Omega) = -n \ln(2\pi) - \frac{n}{2} \ln |\Omega| - \frac{1}{2} \text{tr}(\Omega^{-1} V' V).$$

evaluated at the constrained maximum likelihood estimates.

The fact that the  $LM_2$  test (which is asymptotically similar even with weak instruments) can be interpreted as a score test for the log-likelihood function is somewhat surprising since Zivot, Startz and Nelson (1998) show that their version of the likelihood function score test has poor size properties. The difference arises not from the score itself, but from the estimate used for the variance of the score. The statistic  $LM_2$  uses the asymptotic variance of the score, evaluated at the constrained MLE. The statistic analyzed by Zivot, Startz and Nelson uses instead the Hessian of the concentrated log likelihood function. Although the two tests are asymptotically equivalent, they have different size properties when  $\pi$  is near the origin. In particular, under the weak-instrument asymptotics employed by Staiger and Stock (1997), the Zivot, Startz and Nelson test is not asymptotically similar whereas the  $LM_2$  test is.

Like the Anderson-Rubin statistic, both score tests are approximately pivotal and the critical value function collapses to a constant. However, the conditional approach is also valid for tests based on nonpivotal statistics and nondegenerate critical value functions.

### 1.3 The Wald Test

Likewise, a modified version of the  $W_0$  test for the unknown  $\Omega$  case is based on the statistic

$$(4) \quad W = \frac{(b_{2SLS} - \beta_0)' y_2' N_Z y_2}{\hat{\sigma}^2}.$$

where  $\hat{\sigma}^2 = [1 - b_{2SLS}] \hat{\Omega} [1 - b_{2SLS}]'$ . The critical value function derived for  $W_0$  can then be applied here, but with  $\tilde{T} = Z' Y \hat{\Omega}^{-1} a$  instead of  $T$ . That is, when  $\Omega$  is unknown, one would reject the null hypothesis that  $\beta = \beta_0$  if

$$W - c_W(\tilde{T}, \beta_0, \alpha) > 0.$$

### 1.4 The Likelihood-Ratio Test

Substituting  $\Omega$  for  $\hat{\Omega}$  in the likelihood ratio statistic:

$$LR_1 = \lambda^{\max} - \frac{a' \hat{\Omega}^{-1} Y' N_Z Y \hat{\Omega}^{-1} a}{a' \hat{\Omega}^{-1} a}$$

where  $\lambda^{\max}$  is the largest eigenvalue of  $\hat{\Omega}^{-1/2} Y' N_Z Y \hat{\Omega}^{-1/2}$ . Alternatively, one could use the actual likelihood ratio statistic for the normal distribution with unknown variance,

$$LR = \frac{n}{2} \ln \left( 1 + \frac{b_0' Y' N_Z Y b_0}{b_0' Y' M_X Y b_0} \right) - \frac{n}{2} \ln \left( 1 + \frac{\hat{\lambda}^{\min}}{n - k} \right).$$

Even for relatively small samples, the  $LR_1$  and  $LR$  statistics are close to the  $LR_0$  statistic. Therefore, the critical values in Table 1 can be used for the conditional  $LR_1$  and  $LR$  tests as well, after replacing  $\Omega$  by  $\hat{\Omega}$  in the expression for  $\bar{T}$ .

## 2 Weak Instrument Asymptotics

To examine the approximate properties of test statistics and estimators in models where identification is weak, Staiger and Stock (1997) considered “weak-instrument” asymptotics where the matrix  $\Pi$  converges to the zero matrix as the sample size  $n$  increases. Using this approach, we find that, under some regularity conditions, the limiting rejection probabilities of our conditional tests based on an estimated  $\Omega$  equal the exact rejection probabilities when the errors are normal with known variance. Thus, as long as some regularity conditions are respected, the actual power functions for normal errors with known variance can also be interpreted as the asymptotic power function for nonnormal errors with unknown variance. In particular, Theorem 1 extends Staiger and Stock (1997), Wang and Zivot (1998) and Kleibergen (2001), employing weak-instrument asymptotics to study not only size but also power properties of many tests.

**Theorem 1** *Consider the structural model given by equations (1) and (2) in Chapter*

*1. Suppose*

$$i. Z'Z/n \xrightarrow{p} Q \text{ where } Q \text{ is positive definite and } Z'V/\sqrt{n} \xrightarrow{d} \Psi_{zv} \text{ where } \text{vec}(\Psi_{zv}) \rightsquigarrow$$

$N(0, \Omega \otimes Q)$ .

ii.  $\Pi = C/\sqrt{n}$ , where  $C$  is a fixed  $k$ -dimensional vector.

iii.  $\bar{\psi}$  is continuous function that satisfies the homogeneity condition

$$\bar{\psi}(S, \hat{T}, Z'Z, \hat{\Omega}, \beta_0) = \bar{\psi}(n^{-1/2}S, n^{-1/2}\hat{T}, n^{-1}Z'Z, \hat{\Omega}, \beta_0).$$

iv. The critical value function  $c_{\bar{\psi}}$  derived under the assumption of normality and known  $\Omega$  is continuous.

Then the conditional test based on the statistic  $\bar{\psi}(S, \hat{T}, Z'Z, \hat{\Omega}, \beta_0)$  has limiting rejection probability equal to the exact rejection probability of the conditional test based on  $\bar{\psi}(S, T, nQ, \Omega, \beta_0)$  derived under the assumption of normality with known variance.

**Proof.** By definition,

$$\begin{aligned} \frac{1}{\sqrt{n}} [S, T] &= \frac{1}{\sqrt{n}} Z' [Z\Pi A'b_0 + Vb_0, Z\Pi A'\Omega^{-1}A_0 + V\Omega^{-1}A_0] \\ &= \frac{1}{\sqrt{n}} Z'Z\Pi [A'b_0, A'\Omega^{-1}A_0] + \frac{1}{\sqrt{n}} Z'V [b_0, \Omega^{-1}A_0] \end{aligned}$$

Under Assumption (i),

$$\frac{1}{\sqrt{n}} Z'Z\Pi [A'b_0, A'\Omega^{-1}A_0] \xrightarrow{p} QC [\beta - \beta_0, A'\Omega^{-1}A_0]$$

using the fact that  $A'b_0 = \beta - \beta_0$ . Let  $u \equiv Vb_0$  and  $\varepsilon \equiv V\Omega^{-1}A_0$ . Then, we have:

$$\frac{1}{\sqrt{n}}Z'V [b_0, \Omega^{-1}A_0] \xrightarrow{d} [\Psi_{zu}, \Psi_{z\varepsilon}]$$

where  $\Psi_{zu} \equiv \Psi_{zv}b_0$  and  $\Psi_{z\varepsilon} \equiv \Psi_{zv}\Omega^{-1}A_0$ . In particular, notice that  $\Psi_{zu}$  is independent of  $\Psi_{z\varepsilon}$  since  $u$  is uncorrelated with  $\varepsilon$ . The statistic  $T$  is a function of the unknown variance of the disturbances. However,

$$\frac{1}{\sqrt{n}}(\hat{T} - T) = \frac{1}{\sqrt{n}}Z'Y \left\{ \hat{\Omega}^{-1} - \Omega^{-1} \right\} A_0 \xrightarrow{p} 0$$

since  $Z'Y/\sqrt{n}$  converges in distribution and  $\hat{\Omega}^{-1} - \Omega^{-1} \xrightarrow{p} 0$ . Therefore,  $\bar{\psi}$  has the same limiting distribution as

$$(5) \quad \bar{\psi}(\Psi_{zu} + QC(\beta - \beta_0), \Psi_{z\varepsilon} + QCA'\Omega^{-1}A_0, Q, \Omega, \beta_0)$$

using the Assumption (iii). Analogously, using Assumptions (iii) and (iv) the critical value function  $c_{\bar{\psi}}$  converges in distribution to:

$$(6) \quad c_{\bar{\psi}}(\Psi_{z\varepsilon} + QCA'\Omega^{-1}A_0, Q, \Omega, \beta_0).$$

Consequently,  $\bar{\psi}(S, T, \Omega, \beta_0) - c_{\bar{\psi}}(T, \Omega, \beta_0, \alpha)$  converges in distribution to the difference in expressions (5) and (6). ■

Assumption (i) is similar to that made in the standard asymptotic theory for instrumental variable estimation. If  $Z$  is nonrandom with bounded elements and the errors are i.i.d. with finite second moment, the first part of assumption (i) implies the second part. When lagged endogenous variables appear in  $Z$ , somewhat stronger conditions are needed. Of course, approximations based on (i) can be poor in small samples if the error distribution has very thick tails.

Assumption (ii) states that the coefficients on the instruments are in a neighborhood of zero. Note that  $C$  is allowed to be the zero vector so that Theorem 1 holds even when the structural parameter is not identified. Finally, Assumptions (iii) and (iv) appear to be satisfied for all the commonly proposed test statistics, including the (conditional) Anderson-Rubin, score, likelihood ratio and Wald tests.

### 3 Monte Carlo Simulations

Theorem 1 shows that, under some regularity conditions, the conditional approach leads to *asymptotically* similar tests even when the errors are nonnormal and the reduced form error covariance matrix has been estimated. In this section, we present some evidence suggesting that the weak instrument asymptotics works quite well in moderate sized samples. To evaluate the rejection probability under  $H_0$ , the design I by Staiger and Stock (1997) is once more replicated. Results are reported for the same parameter values except for sample size.

Tables 2 and 3 present rejection probabilities for the following tests: Anderson-

Rubin<sup>1</sup> ( $AR$ ), the Hessian-based score test ( $LM$ ) described by Zivot, Startz and Nelson (1998), the score test ( $LM_2$ ) described in Chapter 2, the likelihood ratio test ( $LR$ ), the conditional likelihood ratio test ( $LR^*$ ), the Wald test centered around the 2SLS estimator ( $W$ ), and the conditional Wald test ( $W^*$ ). The  $AR$ ,  $LM_2$ ,  $LR^*$ , and  $W^*$  tests are approximately similar, whereas the  $LM$ ,  $LR$ , and  $W$  test are not.

Although the  $LM$  test does not present good size properties, the  $LM_2$  test does. Likewise, the  $LR$  and  $W$  tests present worse size properties than the conditional  $LR^*$  and  $W^*$  tests. The null rejection probabilities of the  $LR$  test range from 0.048-0.220 and those of the  $W$  test range from 0.002-0.992 when the number of observations is 80. The null rejection probabilities of their conditional counterparts range from 0.046-0.075 and 0.030-0.072, respectively.

Results for non-normal disturbances are analogous<sup>2</sup>. Tables 4 and 5 show the rejection probabilities of some 5% tests when Staiger and Stock's design II is used. The structural disturbances,  $u$  and  $v_2$ , are serially uncorrelated with  $u_t = (\xi_{1t}^2 - 1)/\sqrt{2}$  and  $v_{2t} = (\xi_{2t}^2 - 1)/\sqrt{2}$  where  $\xi_{1t}$  and  $\xi_{2t}$  are normal with unit variance and correlation  $\sqrt{\rho}$ . The  $k$  instruments are indicator variables with equal number of observations in each cell. When the number of observations is 80, the rejection probabilities under  $H_0$  of the  $LR^*$  and  $W^*$  tests are still close to 5% for all values of  $\lambda'\lambda/k$  and  $\rho$ .

Finally, Tables 6 and 7 compare the power of the conditional  $LR_0^*$  test ( $\Omega$  known) with that of the conditional  $LR^*$  test ( $\Omega$  unknown) when Staiger and Stock's design

---

<sup>1</sup>For the  $AR$  test, a  $\chi^2(k)$  critical value was used.

<sup>2</sup>Once more, the critical value function is based on 1,000 Monte Carlo simulations as if the disturbances were normally distributed with known variance  $\Omega$ .

I with 100 observations is used. The difference between the two power curves is small, which suggests that the power comparison in Chapter 3 for the  $LR_0^*$  test is also valid for the  $LR^*$  test. Tables 8 and 9 show that the same conclusion holds for the conditional  $W_0^*$  and  $W^*$  tests.

## 4 Confidence Regions

Confidence regions for  $\beta$  with approximately correct coverage probability can be constructed by inverting approximately similar tests. Although Dufour (1997) showed that Wald-type confidence intervals are not valid when identification can be arbitrarily weak, the confidence regions based on the conditional Wald test have correct coverage probability in large samples no matter how weak the instruments. Likewise, if the score test or the conditional likelihood ratio test are used, the resulting confidence regions have approximately correct level. Moreover, the regions based on the conditional Wald test necessarily contain the 2SLS estimator of  $\beta$  while the ones based on the conditional likelihood ratio test or on the score test are centered around the limited-information maximum likelihood estimator of  $\beta$ . Therefore, confidence regions based on these tests can be used as evidence of the accuracy of their respective estimators.

To illustrate how informative the confidence regions based on the conditional likelihood ratio test are when compared to the ones based on the score test, design I of Staiger and Stock (1997) is once more used. One sample was drawn where the

true value of  $\beta$  was zero. Figures 4-6 plot the likelihood ratio and score statistics and their respective critical value functions for different values of  $\rho$  at significance level of 5% against  $\beta_0^3$ . The region in which each statistic is below its critical value curve is the corresponding confidence set.

Figures 4-6 suggests that the conditional *LR* confidence regions are considerably smaller than the *LM* ones, as a result of better power properties of the conditional likelihood ratio test. Even when identification is good, the score test fails to reject some non-local yet relevant alternatives exactly when the direction of  $\pi$  is not estimated accurately by  $\hat{\pi}$ .

## 5 Extensions

The previous theory can easily be extended to a structural equation with more than two endogenous variables and with additional exogenous variables as long as inference is to be conducted on *all* the endogenous coefficients. Consider the structural equation

$$y_1 = Y_2\beta + X\gamma + u$$

where  $Y_2$  is the  $n \times l$  matrix of observations on the  $l$  explanatory endogenous variable and  $X$  is the  $n \times r$  matrix of observations on  $r$  exogenous variables. This equation is part of a larger linear system containing the additional exogenous variables  $Z$ . The

---

<sup>3</sup>The approximate critical value function  $\bar{c}(t) = 3.84 + 5.65 \exp(-\bar{t}t/7)$  was used.

reduced form for  $Y = [y_1, Y_2]$  is

$$y_1 = Z\Pi\beta + X\delta + v_1$$

$$Y_2 = Z\Pi + X\Gamma + V_2$$

where  $\delta = \Gamma\beta + \gamma$ . The rows of  $V = [v_1, V_2]$  are i.i.d. normal random vectors with mean zero and covariance matrix  $\Omega$ . It is assumed that  $X$  and  $Z$  have full column rank. The problem is to test the vector hypothesis  $H_0 : \beta = \beta_0$  treating  $\Pi, \Gamma, \delta$  as nuisance parameters.

The unknown parameters associated with  $X$  can be eliminated by taking orthogonal projections. Define the  $l+1$  component column vector  $b = (1, -\beta_0)'$ . Let  $A$  be any  $(l+1) \times l$  matrix whose columns are orthogonal to  $b$ . Then, if  $M_X = I - X(X'X)^{-1}X'$ , the statistics

$$\mathcal{S} = Z'M_X Y b \quad \text{and} \quad \mathcal{T} = Z'M_X Y \Omega^{-1} A$$

are independent and normally distributed. For a nonpivotal statistic  $\psi(\mathcal{S}, \mathcal{T}, \Omega, \beta_0)$ , the critical value can be found computing the  $1 - \alpha$  quantile of the distribution of  $\psi$  conditioned on  $\mathcal{T} = t$ . Again,  $\Omega$  can be replaced with a consistent estimate and the normality assumption dropped without affecting the results under Staiger and Stock asymptotics.

The statistic  $Z'M_X Y$  is a sufficient statistic for the projected data  $M_X Y$ . Furthermore,  $\mathcal{T}$  is a complete sufficient statistic for  $\Pi$  when the null hypothesis is true. Hence

the *Similarity Condition* of Chapter 2 remains valid as long as we restrict attention to tests that are invariant to the nuisance parameters involving  $X$ .

## 6 Conclusions

Replacing the reduced-form covariance matrix with an estimate appears to have little effect on size and power. Even with nonnormal errors, the proposed conditional (pseudo) likelihood ratio test has correct size when identification is weak and good power when identification is strong. This test is asymptotically equivalent to the usual likelihood ratio test under the usual asymptotics. Moreover, power comparison using weak-instrument asymptotics suggests that this test dominates other asymptotically similar tests such as the Anderson-Rubin test and the score test.

Like the Anderson-Rubin and the score approaches, the conditional tests proposed here attain similarity under arbitrarily weak identifiability only when all the unknown endogenous coefficients are being tested. Inference on the coefficient on one endogenous variable when the structural equation contains additional endogenous explanatory variables is not allowed. Duffour (1997) shows how this limitation can be overcome in the context of the Anderson-Rubin test and the same projection approach could presumably be applied here. However, this may entail considerable loss of power.

Finally, the conditional Wald and likelihood ratio tests can also be used to construct confidence regions centered around the 2SLS and LIML estimators, respec-

tively, that have correct coverage probability even when instruments are weak and that are informative when instruments are good.

TABLE 1

## CRITICAL VALUE FUNCTION OF THE LIKELIHOOD RATIO TEST

	$k$							
$\bar{t}'\bar{t}$	1	2	3	4	5	6	7	8
0	3.96	5.88	7.79	9.45	11.19	12.44	14.20	15.74
1	3.80	5.64	7.13	8.67	10.25	11.73	13.09	14.73
2	3.79	5.12	6.62	8.04	9.40	10.74	12.36	13.72
3	3.73	4.94	6.46	7.60	8.67	10.12	11.69	13.24
4	3.85	4.81	5.72	6.93	8.31	9.71	10.61	12.34
5	3.85	4.48	5.45	6.49	7.54	8.77	10.17	11.37
6	3.76	4.23	5.12	6.16	7.25	8.26	9.42	10.66
7	3.78	4.43	5.00	5.84	6.65	7.88	8.96	9.89
8	3.89	4.19	4.96	5.63	6.40	7.32	8.54	9.46
9	3.91	4.19	5.00	5.39	6.08	7.01	7.78	8.96
10	3.83	4.27	4.64	5.09	5.79	6.72	7.45	8.18
15	3.78	4.06	4.36	4.67	5.01	5.45	6.06	6.67
20	3.88	3.97	4.21	4.38	4.74	4.93	5.25	5.61
25	3.79	4.12	4.22	4.29	4.47	4.79	4.96	5.17
30	3.91	3.96	4.00	4.25	4.28	4.69	4.83	4.99
40	3.93	3.93	3.98	4.20	4.34	4.26	4.55	4.72
50	3.85	3.91	4.01	4.20	4.10	4.35	4.44	4.48
75	3.73	3.85	4.02	3.88	4.04	4.14	4.13	4.05
100	3.86	3.74	3.79	3.93	4.06	4.07	4.12	4.04
150	3.74	3.92	3.96	3.88	4.06	3.92	3.99	4.03
200	3.72	3.91	3.93	3.95	3.90	4.03	4.00	4.04
250	3.74	3.78	3.87	4.02	3.86	3.89	4.02	3.96
500	3.72	3.92	3.79	3.94	3.93	3.99	3.89	4.01
750	3.86	3.74	3.64	3.89	3.86	3.82	3.77	3.86
1000	3.87	3.90	3.82	3.94	3.84	3.94	3.68	3.87
5000	3.86	3.90	3.94	3.81	3.84	3.83	3.82	3.76
10000	3.76	3.79	3.78	3.86	3.82	3.88	3.88	3.78
50000	3.88	3.90	3.92	3.71	3.88	3.82	3.94	3.72

TABLE 1 (CONTINUED)

CRITICAL VALUE FUNCTION OF THE LIKELIHOOD RATIO TEST								
	$k$							
$\bar{t}t$	9	10	15	20	25	50	100	200
0	16.85	18.26	24.95	31.23	37.92	67.42	123.85	234.11
1	15.96	17.39	23.98	30.72	36.69	66.18	123.45	232.64
2	15.18	16.57	23.15	29.68	35.83	65.89	122.33	232.26
3	14.40	16.04	22.35	28.40	34.84	64.04	121.60	230.63
4	13.59	14.77	21.22	27.27	33.81	63.73	120.42	229.70
5	12.92	13.90	20.57	26.74	33.01	62.35	119.61	228.79
6	11.97	13.34	19.66	26.04	31.88	61.49	118.52	228.29
7	11.48	12.46	18.90	24.66	31.15	60.20	117.61	227.04
8	10.74	11.86	17.58	24.06	30.05	59.49	116.30	226.37
9	9.84	10.99	16.99	22.99	29.18	58.62	115.21	225.48
10	9.30	10.37	15.89	21.87	28.23	58.35	114.48	224.54
15	7.07	7.91	12.52	18.10	23.66	53.46	109.76	218.54
20	5.82	6.41	9.58	14.07	19.38	47.98	104.72	214.16
25	5.53	5.88	7.88	11.35	15.77	43.23	99.43	209.40
30	5.05	5.51	7.09	9.11	12.67	38.70	94.78	204.80
40	4.65	4.86	5.71	7.05	9.00	29.47	84.80	194.81
50	4.60	4.73	5.19	6.15	7.28	21.44	75.25	184.99
75	4.42	4.36	4.56	5.11	5.67	10.13	51.29	159.19
100	4.10	4.22	4.45	4.76	4.90	7.36	30.38	135.05
150	4.10	4.13	4.24	4.47	4.62	5.72	10.68	86.07
200	3.98	3.98	3.97	4.34	4.28	5.03	7.65	42.16
250	3.89	3.92	3.99	4.21	4.29	4.54	6.40	17.79
500	3.92	3.99	3.93	4.10	4.04	4.26	4.78	6.40
750	3.92	3.95	4.02	3.97	3.97	4.02	4.49	5.27
1000	3.84	3.91	3.88	3.91	3.85	4.05	4.35	4.72
5000	3.76	3.73	3.93	3.88	3.94	3.81	4.03	4.10
10000	3.75	3.97	3.89	3.94	3.82	3.86	3.87	3.95
50000	3.85	3.74	3.84	3.84	3.84	3.88	3.80	3.77

TABLE 2

PERCENT REJECTED UNDER  $H_0$  AT NOMINAL LEVEL OF 5% (20 OBS.)

$\rho$	$\lambda\lambda/k$	$AR$	$LM$	$LM_2$	$LR$	$LR^*$	$W$	$W^*$
0.00	0.00	9.50	4.20	6.70	26.40	9.70	0.80	4.40
0.00	1.00	9.40	3.80	5.00	16.90	8.90	2.10	4.10
0.00	10.00	9.60	5.70	5.60	8.50	7.30	6.50	7.80
0.50	0.00	9.10	14.40	6.70	26.10	10.80	16.20	8.20
0.50	1.00	8.30	8.20	6.20	17.70	8.30	13.40	6.20
0.50	10.00	9.10	4.00	4.90	7.10	6.20	7.50	4.70
0.99	0.00	9.80	46.60	6.60	25.60	11.10	98.40	9.10
0.99	1.00	9.70	22.40	4.50	6.30	6.30	66.50	10.30
0.99	10.00	9.10	5.90	4.20	5.10	5.50	13.10	6.00

TABLE 3

PERCENT REJECTED UNDER  $H_0$  AT NOMINAL LEVEL OF 5% (80 OBS.)

$\rho$	$\lambda\lambda/k$	$AR$	$LM$	$LM_2$	$LR$	$LR^*$	$W$	$W^*$
0.00	0.00	6.20	4.00	5.30	20.70	5.00	0.20	3.00
0.00	1.00	5.30	4.90	5.50	16.20	6.30	1.00	5.20
0.00	10.00	6.10	4.60	4.60	5.80	4.60	3.30	4.00
0.50	0.00	6.70	13.00	5.30	22.00	5.60	13.00	5.10
0.50	1.00	6.10	9.00	5.00	13.80	5.60	12.30	6.10
0.50	10.00	6.10	4.20	4.40	4.80	4.60	5.10	4.00
0.99	0.00	7.30	41.60	5.80	21.80	7.50	99.20	7.20
0.99	1.00	6.50	22.00	4.80	5.00	4.80	60.50	7.00
0.99	10.00	6.40	6.60	5.50	6.00	6.10	13.40	5.80

TABLE 4

PERCENT REJECTED UNDER  $H_0$  AT NOMINAL LEVEL OF 5% (20 OBS.)

- NON-NORMAL DISTURBANCES AND BINARY INSTRUMENTS -

$\rho$	$\lambda'\lambda/k$	$AR$	$LM$	$LM_2$	$LR$	$LR^*$	$W$	$W^*$
0.00	0.00	12.60	6.40	8.40	27.60	13.60	2.40	5.40
0.00	1.00	10.70	5.40	6.80	25.70	10.90	2.10	5.00
0.00	10.00	11.30	8.70	8.50	13.60	11.30	6.90	8.30
0.50	0.00	10.00	8.70	8.30	25.30	12.40	5.60	5.50
0.50	1.00	9.10	6.30	6.80	22.70	10.30	5.50	3.90
0.50	10.00	9.10	7.90	7.60	12.70	9.30	7.40	7.70
0.99	0.00	14.30	48.60	10.70	30.70	15.50	97.50	12.80
0.99	1.00	10.10	27.90	10.20	12.30	11.90	81.00	7.50
0.99	10.00	12.30	12.50	9.50	11.00	10.90	27.60	4.70

TABLE 5

PERCENT REJECTED UNDER  $H_0$  AT NOMINAL LEVEL OF 5% (80 OBS.)

- NON-NORMAL DISTURBANCES AND BINARY INSTRUMENTS -

$\rho$	$\lambda'\lambda/k$	$AR$	$LM$	$LM_2$	$LR$	$LR^*$	$W$	$W^*$
0.00	0.00	6.20	4.40	5.30	23.80	5.90	0.30	3.80
0.00	1.00	6.40	4.00	5.50	22.50	6.50	0.20	3.80
0.00	10.00	5.90	7.30	7.90	12.10	8.10	2.90	7.10
0.50	0.00	7.20	8.60	6.60	23.40	7.90	4.40	5.60
0.50	1.00	6.50	6.70	6.30	21.80	7.50	3.10	5.40
0.50	10.00	6.70	6.70	6.70	10.80	7.40	4.20	5.40
0.99	0.00	7.60	41.30	7.10	24.30	7.90	96.90	7.00
0.99	1.00	6.60	29.20	6.70	8.40	7.00	81.20	5.60
0.99	10.00	5.70	11.10	6.40	6.70	7.20	27.40	3.10

TABLE 6

PERCENT REJECTED AT NOMINAL LEVEL OF 5%

CONDITIONAL LIKELIHOOD RATIO TEST - WEAK INSTRUMENTS

$\beta$	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	$LR_0^*$	$LR^*$	$LR_0^*$	$LR^*$	$LR_0^*$	$LR^*$
-10.00	30.10	31.70	34.20	35.10	56.90	57.70
-8.00	28.20	29.80	34.00	34.10	60.30	60.40
-6.00	28.90	29.60	34.70	34.90	62.30	61.80
-4.00	28.00	29.30	35.90	36.90	71.50	71.70
-2.00	24.00	24.80	37.80	38.00	96.70	96.10
0.00	5.40	5.70	5.90	6.40	6.30	7.00
2.00	26.60	26.00	21.70	23.30	24.90	26.10
4.00	27.10	28.00	25.60	26.00	33.70	33.80
6.00	29.00	31.40	27.60	28.90	37.00	37.20
8.00	30.50	30.90	30.20	29.70	41.00	41.60
10.00	28.80	30.50	28.20	30.10	43.20	45.00

TABLE 7

PERCENT REJECTED AT NOMINAL LEVEL OF 5%

CONDITIONAL LIKELIHOOD RATIO TEST - GOOD INSTRUMENTS

$\beta$	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	$LR_0^*$	$LR^*$	$LR_0^*$	$LR^*$	$LR_0^*$	$LR^*$
-10.00	99.80	99.80	100.00	100.00	100.00	100.00
-8.00	99.70	99.80	100.00	100.00	100.00	100.00
-6.00	98.90	98.70	100.00	100.00	100.00	100.00
-4.00	95.10	95.10	100.00	99.90	100.00	100.00
-2.00	58.50	59.00	78.90	78.90	98.40	98.60
0.00	5.40	5.90	5.30	5.50	6.40	7.00
2.00	59.70	60.00	48.90	48.60	41.70	42.90
4.00	94.30	93.70	85.10	84.40	78.60	78.70
6.00	98.80	98.70	96.00	95.60	90.80	90.60
8.00	99.40	99.10	97.50	97.00	95.70	94.70
10.00	99.80	99.80	99.50	99.30	95.50	95.50

TABLE 8

PERCENT REJECTED AT NOMINAL LEVEL OF 5%

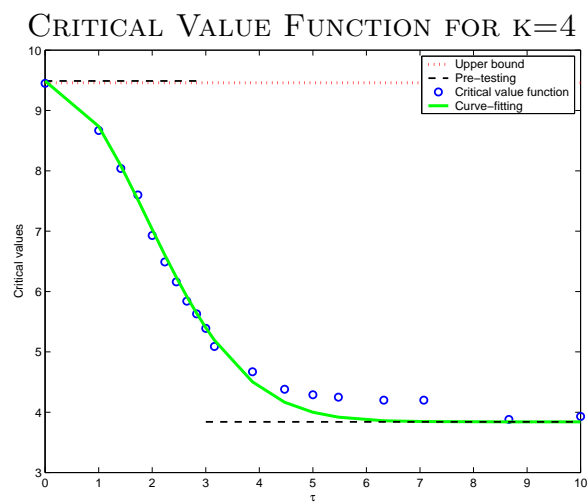
CONDITIONAL WALD TEST - WEAK INSTRUMENTS						
$\beta$	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	$W_0^*$	$W^*$	$W_0^*$	$W^*$	$W_0^*$	$W^*$
-10.00	29.70	30.70	30.40	30.80	2.00	2.80
-8.00	30.70	31.70	31.30	32.80	2.00	2.60
-6.00	29.80	31.40	32.30	32.70	2.50	3.30
-4.00	29.20	29.70	32.30	33.00	2.40	2.40
-2.00	25.20	26.50	33.60	35.60	25.40	27.70
0.00	4.90	4.70	3.60	4.50	3.70	5.20
2.00	27.50	28.00	22.60	22.50	0.90	1.10
4.00	28.00	30.60	23.70	24.60	1.30	1.90
6.00	29.30	29.10	25.80	25.80	1.90	1.90
8.00	31.80	32.20	29.10	29.10	1.80	2.10
10.00	29.10	30.20	26.30	27.50	1.70	1.80

TABLE 9

PERCENT REJECTED AT NOMINAL LEVEL OF 5%

CONDITIONAL WALD TEST - GOOD INSTRUMENTS						
$\beta$	$\rho = 0.00$		$\rho = 0.50$		$\rho = 0.99$	
	$W_0^*$	$W^*$	$W_0^*$	$W^*$	$W_0^*$	$W^*$
-10.00	99.80	99.90	100.00	100.00	100.00	100.00
-8.00	100.00	100.00	100.00	100.00	100.00	100.00
-6.00	99.70	99.40	100.00	100.00	100.00	100.00
-4.00	97.30	97.40	100.00	100.00	100.00	100.00
-2.00	70.20	69.90	73.60	74.10	43.70	43.80
0.00	5.50	5.60	4.30	4.50	5.00	5.40
2.00	70.10	70.20	58.80	59.40	48.00	47.80
4.00	96.20	95.90	87.30	86.80	75.10	73.70
6.00	99.20	99.00	94.80	93.70	83.90	82.20
8.00	99.70	99.40	96.40	95.30	85.60	83.30
10.00	99.80	99.70	97.20	96.60	80.00	78.90

FIGURE 1



Student Version of MATLAB

FIGURE 2

## EMPIRICAL POWER OF TESTS: WEAK INSTRUMENTS

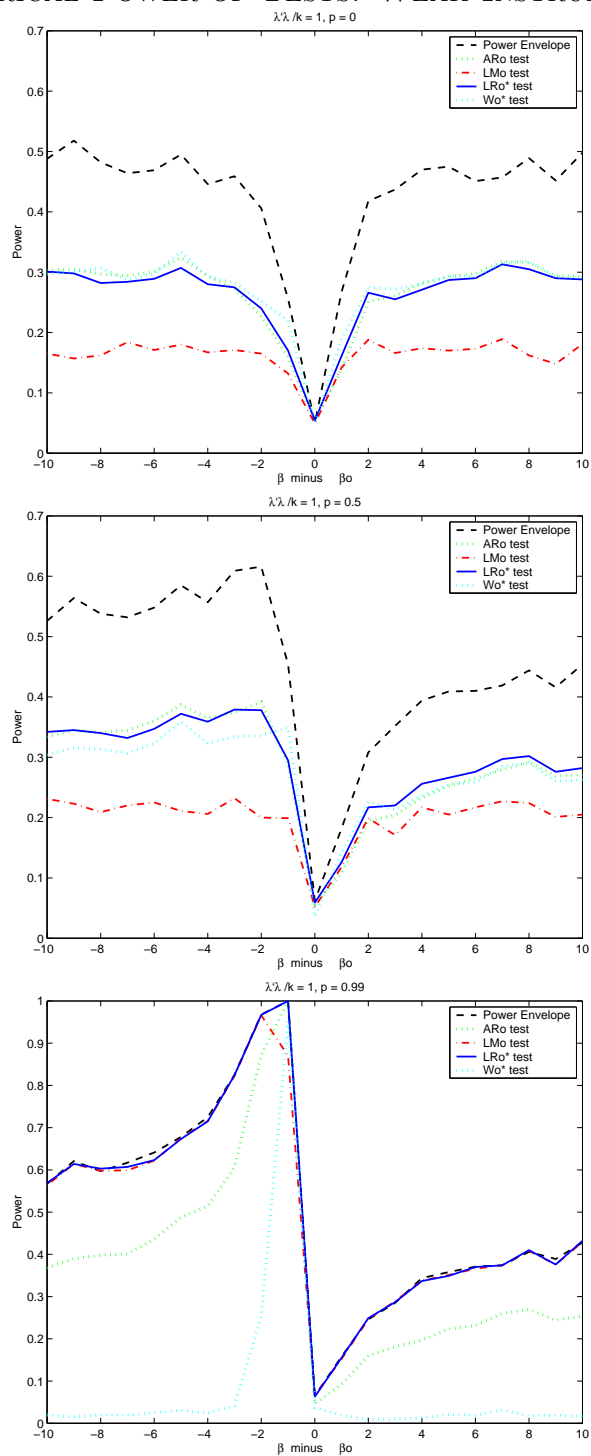


FIGURE 3

## EMPIRICAL POWER OF TESTS: GOOD INSTRUMENTS

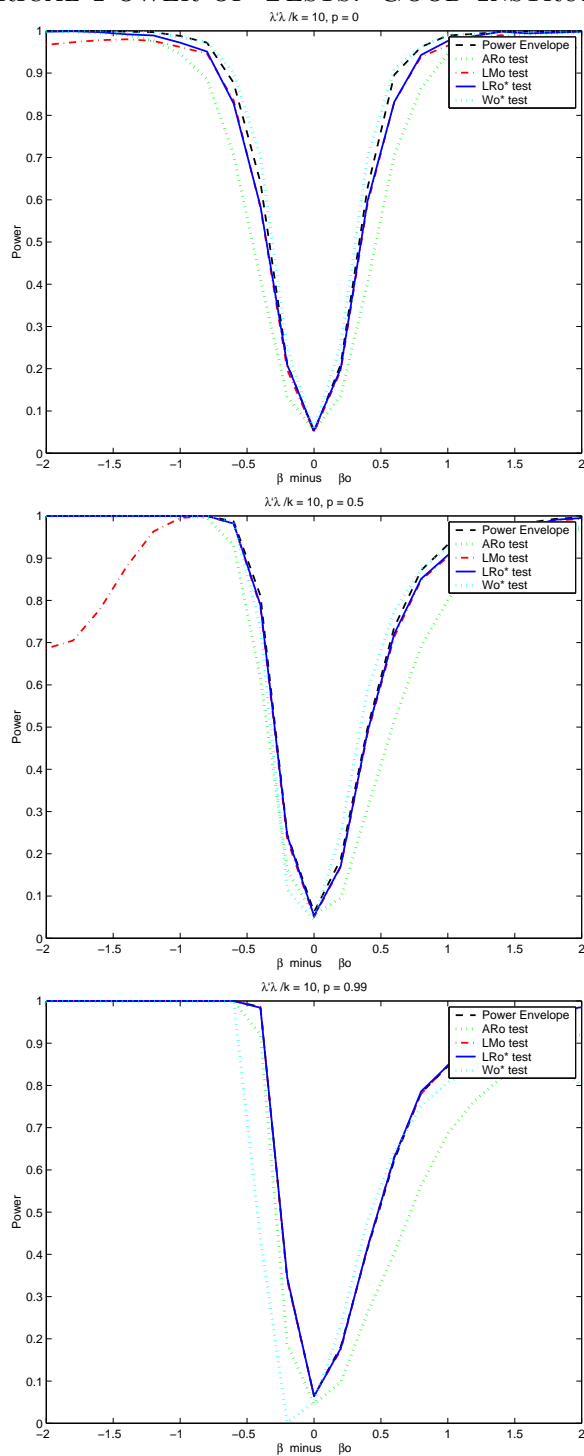


FIGURE 4

## CONFIDENCE REGIONS: INVALID INSTRUMENTS

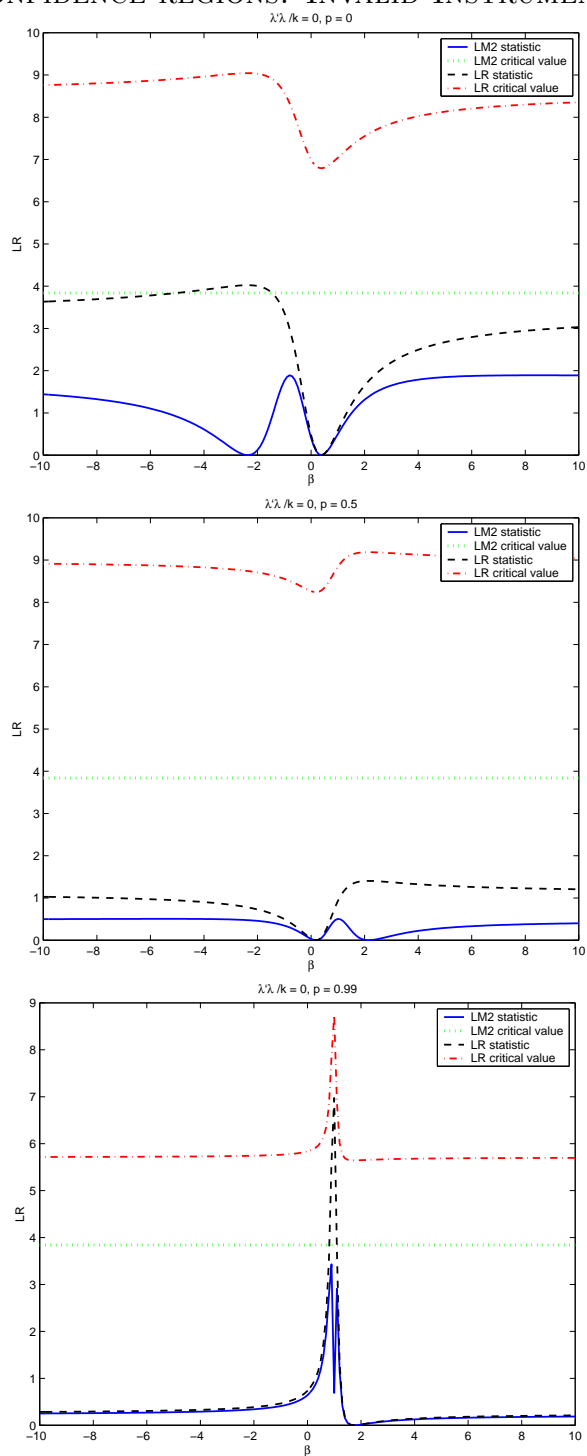


FIGURE 5

## CONFIDENCE REGIONS: WEAK INSTRUMENTS

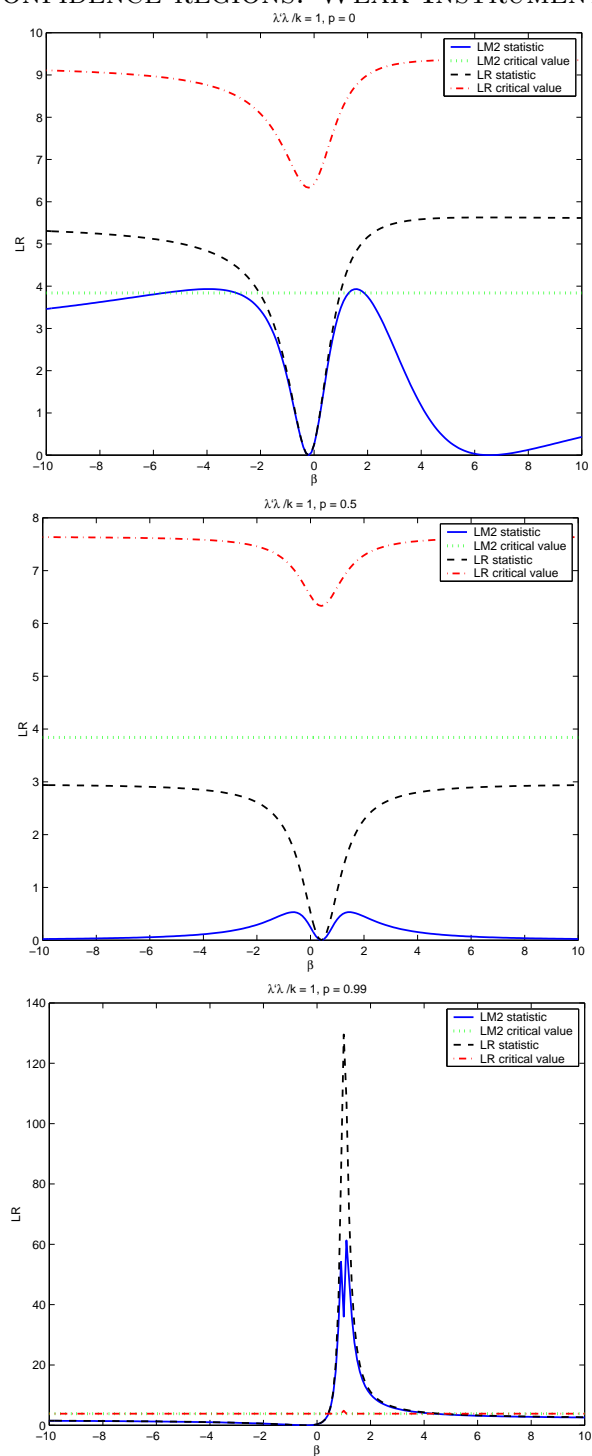
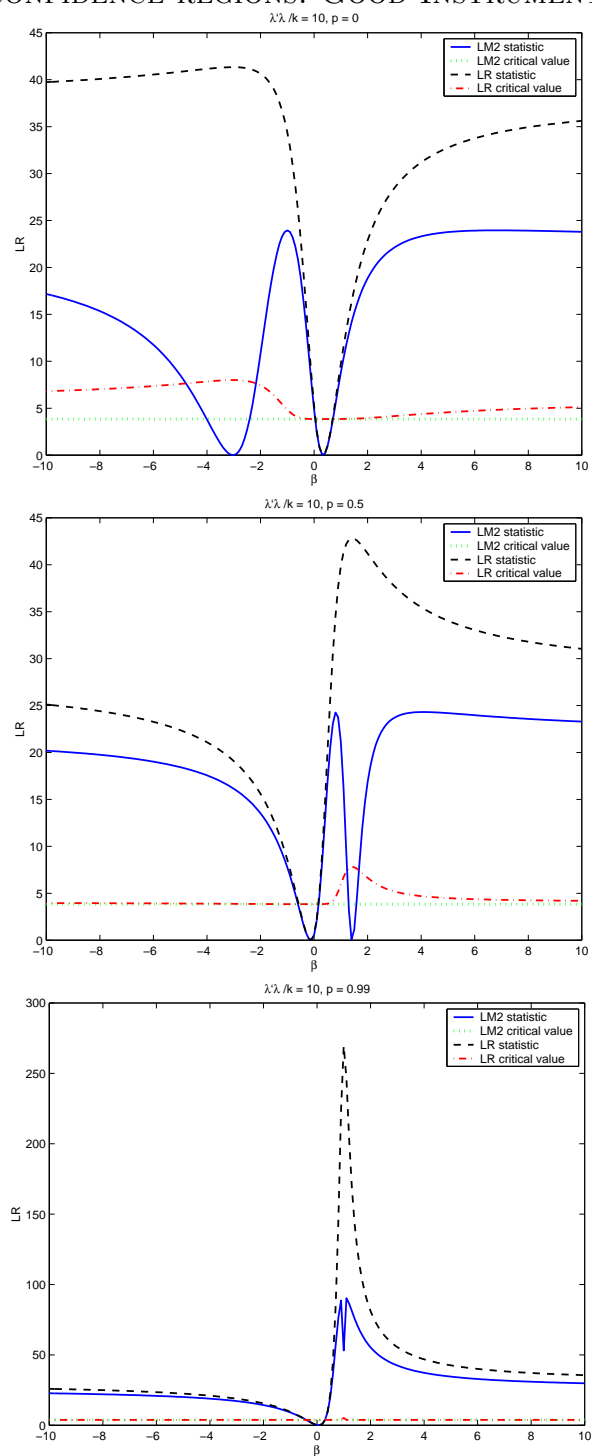


FIGURE 6

## CONFIDENCE REGIONS: GOOD INSTRUMENTS



# Appendix A

## Lemmas and Theorems

The results in Chapter 2 use the following two lemmas proved in Lehmann (1986), pp. 142-3:

LEMMA A.1: *Let  $X$  be a random vector with probability distribution*

$$dP_{\theta}(x) = C(\theta) \exp \left[ \sum_{j=1}^k T_j(x) \theta_j \right] d\mu(x)$$

*and let  $P^T$  be the family of distributions of  $T = (T_1(X), \dots, T_k(X))$  as  $\theta$  ranges over the set  $W$ . Then  $P^T$  is complete provided  $W$  contains a  $k$ -dimensional rectangle.*

LEMMA A.2: *Suppose that the distribution of  $X$  is given by*

$$dP_{\theta, \mathcal{V}}(x) = C(\theta, \mathcal{V}) \exp \left[ \theta R(x) + \sum_{j=1}^k \mathcal{V}_j T_j(x) \right] d\mu(x)$$

*where the  $V_j$  are the nuisance parameters and  $\mu$  is absolutely continuous with respect to the Lebesgue measure. Suppose that  $S = h(R, T)$  is independent of  $T$  when  $\theta = \theta_0$  and that*

$$h(r, t) = a(t)r + b(t) \quad \text{with } a(t) > 0.$$

Then the uniformly most powerful unbiased (UMPU) test  $\phi$  for  $H_0 : \theta = \theta_0$  against  $H_1 : \theta \neq \theta_0$  is given by

$$\phi(s) = \begin{cases} 1 & \text{if } s < C_1 \text{ or } s > C_2 \\ 0 & \text{otherwise} \end{cases}$$

where  $C_1$  and  $C_2$  are determined by  $E_0 \{\phi(S)\} = \alpha$  and  $E_0 \{S\phi(S)\} = \alpha E_0 \{S\}$ .

PROOF OF THEOREM 2.2: The following is true:

a. For some measure  $\mu(y)$ , the probability distribution of  $Y$  can be written as:

$$dP_{\theta, \pi}(y) = C(\theta, \pi) \exp[\theta R(y) + \pi T(y)] d\mu(y)$$

where  $R(Y)$  is the first column of  $Z'Y\Omega^{-1}$  and  $\theta = \pi(\beta - \beta_0)$ . Since  $P$  does not contain the origin and the model is just identified, testing  $H_0 : \beta = \beta_0$  against  $H_1 : \beta \neq \beta_0$  is equivalent to testing  $H_0 : \theta = \theta_0$  against  $H_1 : \theta \neq \theta_0$ . Let

$$\bar{S} = \frac{(Z'Z)^{-1/2} Z'(y_1 - y_2\beta_0)}{\sigma_0}.$$

Notice that  $\bar{S} = \delta_1 R + \delta_2 T$  where

$$\delta_1 = \sigma_0 (Z'Z)^{-1/2} \quad \text{and} \quad \delta_2 = \frac{-\omega_{22}\beta_0 + \omega_{12}}{\sigma_0} (Z'Z)^{-1/2}$$

**Proof.**

Now Lemma A.2 can be applied. Since  $\bar{S} \sim N(0, 1)$  under  $H_0$  and, in particular, it is symmetric around zero, it is straightforward to show that the optimal test rejects the null if  $AR_0 > c_\alpha$ . Under the alternative  $\beta$ ,

$$AR_0 \sim \chi^2 \left( 1, \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\sigma_0^2} \right).$$

Consequently, the power of the optimal test is given by (6).

b. Since  $\pi$  is known, for some measure  $\mu(y)$ , the probability distribution can be written as:

$$dP_{\beta, \pi}(y) = C(\beta, \pi) \exp [R(y)' \pi \beta] d\mu(y)$$

Since this distribution is a one-parameter exponential family, the *UMPU* test rejects the null hypothesis if

$$(A.3) \quad \mathcal{R} = \frac{\{\pi' Z' [(y_1 - Z\pi\beta_0) - \omega_{12}\omega_{22}^{-1}(y_2 - Z\pi)]\}^2}{(\omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{12}) \pi' Z' Z \pi}$$

is larger than  $c_\alpha$ . Under the alternative  $\beta$ ,

$$\mathcal{R} \sim \chi^2 \left( 1, \frac{\pi' Z' Z \pi}{(\omega_{11} - \omega_{12}\omega_{22}^{-1}\omega_{12})} (\beta - \beta_0)^2 \right)$$

Consequently, the power of the optimal test is given by (7).

c. The power of the test  $\phi$  is given by  $E_{\beta,\pi}\phi(S,T)$ . Since  $S$  and  $T$  are independent, then:

$$E_{\beta,\pi}\phi(S,T) = \int \left[ \int \phi(s,t) f(s,\beta,\pi) ds \right] g(t,\beta,\pi) dt$$

where  $f(s,\beta,\pi)$  and  $g(t,\beta,\pi)$  are the density functions associated to  $S$  and  $T$ , respectively. Notice that the power conditioned on  $T = t$  is

$$\int \phi(s,t) f(s,\beta,\pi) ds.$$

Consider the test  $\phi^*(S)$  that assigns 1 if  $f(s,\beta,\pi) > kf(s,\beta_0)$  and 0 otherwise, where  $k$  is chosen such that  $E_{\beta_0,\pi}\phi^*(S) = \alpha$ . The claim is that the test  $\phi^*(S)$  is most powerful among all similar tests at the significance level  $\alpha$ .

Let  $S^+$  and  $S^-$  be the sets in the sample space where  $\phi^*(s) - \phi(s,t) > 0$  and  $\phi^*(s) - \phi(s,t) < 0$ , respectively. Notice that, if  $s$  is in  $S^+$ ,  $\phi^*(s) = 1$  and  $f(s,\beta,\pi) > kf(s,\beta_0)$ . Analogously, if  $s$  is in  $S^-$ ,  $\phi^*(s) = 0$  and  $f(s,\beta,\pi) \leq kf(s,\beta_0)$ . Therefore:

$$\int [\phi^*(s) - \phi(s,t)] [f(s,\beta,\pi) - kf(s,\beta_0)] ds \geq 0$$

The difference in power satisfies

$$\int [\phi^*(s) - \phi(s,t)] f(s,\beta,\pi) dv \geq k \int [\phi^*(s) - \phi(s,t)] f(v,\beta_0) ds$$

By Theorem 2.1, if the test  $\phi(S, T)$  is similar then  $E_0\phi(S, t) = \alpha$ , *a.e.*  $\mathcal{P}^T$ . Without loss of generality, it can be considered that  $E_0\phi(S, t) = \alpha$ ,  $\forall t$ . That is:

$$\int \phi(s, t) f(s, \beta, \pi) ds = \alpha \quad , \forall t$$

Therefore, the following holds:

$$\int [\phi^*(s) - \phi(s, t)] f(s, \beta, \pi) ds \geq 0$$

Since the test that maximizes the conditional power does not depend on  $t$ , then this test itself maximizes power, as was to be proved. Since  $S$  is normally distributed,  $f(s, \beta, \pi) > kf(s, \beta_0)$  for some  $k$  such that  $E\phi^*(S) = \alpha$  if and only if the following holds. If  $\beta > \beta_0$  then the test rejects the null if  $\pi'S > z_\alpha \sqrt{\sigma_0^2 \pi' Z' Z \pi}$ . If  $\beta < \beta_0$  then the test rejects the null if  $\pi'S < -z_\alpha \sqrt{\sigma_0^2 \pi' Z' Z \pi}$ , where  $z_\alpha$  is the critical value of a  $\mathcal{N}(0, 1)$  distribution for the significance level  $\alpha$ . For two-sided alternative, the optimal test rejects  $H_0$  if

$$\frac{(\pi'S)^2}{\sigma_0^2 \pi' Z' Z \pi} > c_\alpha$$

Under the alternative  $\beta$ ,

$$\frac{(\pi'S)^2}{\sigma_0^2 \pi' Z' Z \pi} \sim \chi^2 \left( 1, \frac{\pi' Z' Z \pi (\beta - \beta_0)^2}{\sigma_0^2} \right)$$

Consequently, the power envelope is given by (8). ■

## Appendix B

### Likelihood Ratio Derivation

Ignoring an additive constant, the log-likelihood function can be written as:

$$(B.1) \quad l(Y; \beta, \pi, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} \text{tr} (\Omega^{-1} V' V)$$

Maximizing with respect to  $\pi$ , one finds that  $\pi(\beta, \Omega) = (Z'Z)^{-1} Z'Y\Omega^{-1}c/c'\Omega^{-1}c$  where  $c \equiv (\beta, 1)'$ . The concentrated log-likelihood function,  $l_c(Y; \beta, \Omega)$ , defined as  $l(Y; \beta, \pi(\beta, \Omega), \Omega)$ , is given by:

$$l_c(Y; \beta, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} \left[ \text{tr} \Omega^{-1} Y' Y - \frac{c' \Omega^{-1} Y' N Y \Omega^{-1} c}{c' \Omega^{-1} c} \right]$$

When evaluated at  $\hat{\beta}$ , the maximum likelihood estimator of  $\beta$  when  $\Omega$  is known, this becomes

$$l_c(Y; \hat{\beta}, \Omega) = -\frac{n}{2} \ln |\Omega| - \frac{1}{2} [\text{tr} \Omega^{-1} Y' Y - \bar{\lambda}^{\max}]$$

where  $\bar{\lambda}^{\max}$  is the largest eigenvalue of  $(Z'Z)^{-1/2} Z'Y\Omega^{-1}Y'Z (Z'Z)^{-1/2}$  or, equivalently, the largest eigenvalue of  $\Omega^{-1/2} Y' N_Z Y \Omega^{-1/2}$ . Since the likelihood ratio statistic

when  $\Omega$  is known,  $LR_0$ , is defined as  $2 \left[ l_c \left( Y; \hat{\beta}, \Omega \right) - l_c \left( Y; \beta_0, \Omega \right) \right]$ , it follows that:

$$LR_0 = \bar{\lambda}^{\max} - \frac{a' \Omega^{-1} Y' N_Z Y \Omega^{-1} a}{a' \Omega^{-1} a}$$

To find the likelihood ratio when  $\Omega$  is unknown, we maximize (A.1) with respect to  $\Omega$ , obtaining  $\Omega(\beta, \pi) = V'V/n$ . Inserting this into (B.1). we obtain

$$l(Y; \beta, \pi, \Omega(\beta, \pi)) = -\frac{n}{2} \ln |V'V| + n \ln(n) - n$$

After considerable algebra we find that the maximum value of the log-likelihood function for a fixed  $\beta$  is given by:

$$-\frac{n}{2} \ln \left( \frac{u'u}{u'M_Z u} \right) - \frac{n}{2} \ln |Y'M_Z Y| + n \ln(n) - n$$

The concentrated log-likelihood function,  $l_c(Y; \beta)$ , defined as  $l(Y; \beta, \pi(\beta), \Omega(\beta))$ , is then given by:

$$l_c(Y; \beta) = -\frac{n}{2} \ln \left( 1 + \frac{d'Y'N_Z Y d}{d'Y'M_Z Y d} \right) - \frac{n}{2} \ln |Y'M_Z Y| + n \ln(n) - n$$

where  $d = (1, -\beta)'$ . Moreover, the concentrated log-likelihood function evaluated at

the maximum likelihood estimator  $\beta_{LIML}$  is then given by:

$$l_c(Y; \beta_{LIML}) = -\frac{n}{2} \ln \left( 1 + \frac{\lambda^{\min}}{n-k} \right) - \frac{n}{2} \ln |Y' M_Z Y| + n \ln(n) - n$$

Since the  $LR$  when  $\Omega$  is unknown is defined as  $2[l_c(Y; \beta_{LIML}) - l_c(Y; \beta_0)]$ , it follows that:

$$LR = -n \ln \left( 1 + \frac{\lambda^{\min}}{n-k} \right) + n \ln \left( 1 + \frac{b' Y' N_Z Y b}{b' Y' M_Z Y b} \right)$$

## References

ANDREWS, D. AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 62, 1383-414.

ANDERSON, T., N. KUNITOMO AND T. SAWA (1982): "Evaluation of the distribution function of the limited information maximum likelihood estimator," *Econometrica*, 50, 1009-1028.

ANDERSON, T., AND H. RUBIN (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20, 46-63.

ANGRIST, J. AND A. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *The Quarterly Journal of Economics*, 106, 979-1014.

BOUND, J., D. JAEGER AND R. BAKER (1995): "Problems with instrumental variables estimation when the correlation between the Instruments and the endogenous explanatory variables is weak," *Journal of American Statistical Association*, 90, 443-50.

DUFOUR, J-M. (1997): "Some impossibility theorems in econometrics with applications to structural and dynamic models," *Econometrica*, 65, 1365-88.

DUFOUR, J-M. AND J. JASIAK (2001): "Finite sample inference for simultaneous equations and models with unobserved and generated regressors," *International Economic Review*, 42, 815-44.

GLESER, L., AND J. HWANG (1987): "The non-existence of  $100(1-\alpha)\%$  confidence sets of finite expected diameter in errors-in-variables and related models," *Annals of Statistics*, 15, 1351-62.

HAHN, J. AND J. HAUSMAN (2002): "A new specification test for the validity of instrumental variables," *Econometrica*, 70, 163-89.

HOROWITZ, J. (2001): "The bootstrap and hypothesis tests in econometrics," *Journal of Econometrics*, 100, 37-40.

KLEIBERGEN, F. (2001): "Pivotal statistics for testing structural parameters in instrumental variables regression," forthcoming *Econometrica*.

LEHMANN, E. (1986): *Testing Statistical Hypothesis*. 2nd edition, Wiley Series in Probability and Mathematical Statistics.

NELSON, C. AND R. STARTZ (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967-76.

ROTHENBERG, T. (1984): "Approximating the distributions of econometric estimators and test statistics," *Handbook of Econometrics*, ed. by Z. Griliches and M. Intriligator, vol. 2, ch. 15. North-Holland, Amsterdam.

STAIGER, D., AND J. STOCK (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557-86.

VAN GARDEREN, K. J.(2000): "An alternative comparison of classical tests: assessing the effects of curvature" in *Applications of Differential Geometry to Econo-*

*metrics*, ed. by M. Salmon and P. Marriot. Cambridge: Cambridge University Press.

WANG, J., AND E. ZIVOT (1998): "Inference on a structural parameter in instrumental variables regression with weak instruments," *Econometrica*, 66, 1389-404.

ZELLNER, A. (1962): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, pp. 348-68.

ZIVOT, E., R. STARTZ, AND C. NELSON (1998): "Valid confidence intervals and inference in the presence of weak instruments," *International Economic Review*, 39, 1119-44.